

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:24:04 / Search time 1393.27 Seconds  
(without alignments)  
9647.960 Million cell updates/sec

Title: US-09-763-902b-10

Perfect score: 830

Sequence: 1 tgaacttgcatcactg.....ataattaataacacatg 830

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

EST:  
1: em\_estba:\*  
2: em\_estba:\*  
3: em\_estba:\*  
4: em\_estba:\*  
5: em\_estba:\*  
6: em\_estba:\*  
7: em\_estba:\*  
8: em\_estba:\*  
9: em\_estba:\*  
10: em\_estba:\*  
11: em\_estba:\*  
12: em\_estba:\*  
13: em\_estba:\*  
14: em\_estba:\*  
15: em\_estba:\*  
16: em\_estba:\*  
17: em\_estba:\*  
18: em\_estba:\*  
19: em\_estba:\*  
20: em\_estba:\*  
21: em\_estba:\*  
22: em\_estba:\*  
23: em\_estba:\*  
24: em\_estba:\*  
25: em\_estba:\*  
26: em\_estba:\*  
27: em\_estba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	766	92.3	982	9	AL551859
2	678	81.7	814	9	AL575327
3	645	77.7	685	14	BM975482
4	632.8	76.2	1110	13	BM543865
5	631.8	76.1	720	12	BF195575
6	629.4	75.8	631	10	AM274482

Result No.	Score	Match	Query Length	DB ID	Description
7	623	75.1	649	14	BM789812
8	613	73.9	644	14	BO581596
9	603	72.4	906	12	BG109830
10	602	72.5	639	10	BE615929
11	601.4	72.5	1125	14	BO421087
12	592.4	71.4	882	12	BE615173
13	591.8	71.3	882	12	BE794429
14	590.6	71.2	606	10	BE463515
15	589.4	71.0	596	10	BE616032
16	578.4	69.7	1002	13	BM470251
17	560.4	67.5	576	12	BF062685
18	554	66.7	577	14	BM723690
19	553	66.6	553	9	A1739032
20	552	66.5	577	14	BO188558
21	549.6	65.2	973	14	BO889949
22	541.2	65.2	846	10	BE515286
23	540.4	65.1	571	14	BM681875
24	532.8	64.2	545	9	AA694201
25	526	63.4	527	9	A1967926
26	525	63.3	552	14	BM675619
27	508.6	61.3	526	9	A1479307
28	506.4	61.0	534	10	AM204159
29	503.4	60.7	1076	13	BM552305
30	498.8	60.1	942	12	BE790503
31	496.2	59.6	727	13	BI913135
32	495	59.6	511	10	AM205842
33	483	58.2	514	14	BO186422
34	481.8	58.0	868	14	BO937662
35	477.8	57.6	512	12	BG231831
36	476.8	57.4	1033	12	BF683948
37	474	57.1	493	14	BO185616
38	473.4	57.0	1043	12	BG249456
39	471.8	56.8	1057	12	BE733692
40	471	56.7	471	14	BO187897
41	467.8	56.4	786	13	BI560420
42	467.4	56.3	939	11	AK003254
43	463.8	55.9	468	9	A1697839
44	462.4	55.7	478	9	A1807220
45	462.4	55.7	897	14	BO682672

## ALIGNMENTS

RESULT 1  
AL551859  
LOCUS  
DEFINITION  
AL551859 LTI\_NFL006.PL2 Homo sapiens CDNA clone CS0D1061YJ24 5  
PRIME, mRNA sequence.  
ACCESSION  
AL551859  
VERSION  
AL551859.1 GI:128990210  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1. (bases 1 to 982)  
AUTHORS  
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
TITLE  
Full-length CDNA libraries and normalization  
JOURNAL  
unpublished (2001)  
COMMENT  
Contact: Centre National de Sequencage  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seque@genoscope.cns.fr, Web: www.genoscope.cns.fr.  
FEATURES  
source  
1. 982  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0D1061YJ24"  
/clone\_lib="LTI\_NFL006.PL2"  
/issue\_type="placenta"  
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand CDNA  
was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 211 a 270 c 263 g 232 t 6 others  
ORIGIN

Query Match 92.3% Score 766; DB 9; Length 982;  
Best Local Similarity 99.7% Pred. No. 9,2e-202;  
Matches 766; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

49 CCCCTGTTCCCAAGGAGAGAAATACCTGTGAGCCCTCTTCATGAAACCA 108  
215 CCCCTGTTCCCAAGGAGAGAAATACCTGTGAGCCCTCTTCATGAAACCA 274  
109 GATGCAATCTGGGATTTCAGACCTATGTGATCAGGCGCTGAGAGGAGTT 168  
275 GATGCAATCTGGGATTTCAGACCTATGTGATCAGGCGCTGAGAGGAGTT 334  
169 TGTCAATGCTCTACACACCATGATTAAGCGGCGCTTGTCTCCCGCTGTACAT 228  
335 TGTCAATGCTCTACACACCATGATTAAGCGGCGCTTGTCTCCCGCTGTACAT 394  
229 GGGCAGACGACCCCTGTGTGAAATGCAATGCTGTTTCAAGCAAGAAATCTTGA 288  
395 GGGCAGACGACCCCTGTGTGAAATGCAATGCTGTTTCAAGCAAGAAATCTTGA 454  
289 GTTTTGAATGTGCTTCCAGCGAGTTTCCAAATCAGCGGTGAGTGGCAGCTGT 348  
455 GTTTTGAATGTGCTTCCAGCGAGTTTCCAAATCAGCGGTGAGTGGCAGCTGT 514  
349 TCATGATGAAGCCACACCAAGCCAGCGCTTGTGTGATCTGTGATGATGAA 408  
515 TCATGATGAAGCCACACCAAGCCAGCGCTTGTGTGATCTGTGATGATGAA 574  
409 GTTTGAGGGGAAACAAACAGGGGCTTCAACCACTTCAATCCAGCCAGGCTTC 468  
575 GTTTGAGGGGAAACAAACAGGGGCTTCAACCACTTCAATCCAGCCAGGCTTC 634  
469 ACCCAGACACAGTGTGAAAGATCCCAAGTACTGCTTCCAGGAGTGGGCGAG 528  
635 ACCCAGACACAGTGTGAAAGATCCCAAGTACTGCTTCCAGGAGTGGGCGAG 694  
529 CTAGTGGGGGTGGCAGAGTCTCTTTCCTTCAATTCAGCCCTAGCTGTAGAAATGCA 588  
695 CTAGTGGGGGTGGCAGAGTCTCTTTCCTTCAATTCAGCCCTAGCTGTAGAAATGCA 754  
589 AACCTGACTCTCAAGGATGTGAGGAACACAAAGTCAATTTCTGTGTTGGGAGACCTG 648  
755 AACCTGACTCTCAAGGATGTGAGGAACACAAAGTCAATTTCTGTGTTGGGAGACCTG 814  
649 CAGACTCCTAGTCCAGGTTGAATCTTTTGTGTCAGATTTAGAGAGTCCCTTT 708  
815 CAGACTCCTAGTCCAGGTTGAATCTTTTGTGTCAGATTTAGAGAGTCCCTTT 874  
709 CCTGAATATATATCTGTTGTCATAGTTTCTTTTCAAGTAGTAATCTTTATTTT 768  
875 CCTGAATATATATCTGTTGTCATAGTTTCTTTTCAAGTAGTAATCTTTATTTT 934  
769 CTACTTGGCCAGTAGAGCTGTGATCTGGGAATCTGACAAATTAAT 816  
935 CTACTTGGCCAGTAGAGCTGTGATCTGGGAATCTGACAAATTAAT 982

RESULT 2 814 bp. mRNA linear EST 16-FEB-2001  
AL575327/c  
LOCUS AL575327 LFI.NFL006.PL2 Homo sapiens cDNA clone CS0D1061YJ24 3  
DEFINITION prime, mRNA sequence.

ACCESSION AL575327  
VERSION AL575327.1 GI:12936386  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. (bases 1 to 814).  
AUTHORS Li, W.B., Gruber, C., Jesssee, J., and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

# FEATURES

source

1. 814  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0D1061YJ24"  
/clone\_1ib="LFI.NFL006.PL2"  
/tissue\_type="Placenta"  
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com"

BASE COUNT 190 a 219 c 215 g 174 t 16 others  
ORIGIN

Query Match 81.7% Score 678; DB 9; Length 814;  
Best Local Similarity 96.3% Pred. No. 2.4e-177;  
Matches 697; Conservative 15; Mismatches 10; Indels 2; Gaps 2;

50 CCCTGTTCCCAAGGAGAGAAATACCTGTGAGCCCTCTTCATGAAACCA 109  
723 CCCTGTTCCCAAGGAGAGAAATACCTGTGAGCCCTCTTCATGAAACCA 664  
110 ATGCAATCTGTGATTTCAAGACCTATGTGATCAGGCTGAGAGGAGTTT 169  
663 ATGCAATCTGTGATTTCAAGACCTATGTGATCAGGCTGAGAGGAGTTT 604  
170 GTCAATGTCTACTACACACCATGAGTAAGCGGCGGTTTGTCTCCGCTGTACATG 229  
603 GTCAATGTCTACTACACACCATGAGTAAGCGGCGGTTTGTCTCCGCTGTACATG 545  
230 GGCACAGCACCCCTGTGTGAATGCAATGCTGTTTCAAGCAAGAAATCTTGA 289  
544 GGCACAGCACCCCTGTGTGAATGCAATGCTGTTTCAAGCAAGAAATCTTGA 485  
290 TTTTGAATGTGCTTCCAGGAGTTCGAATTCAGCGGTGAGAGTGGCAGCTGT 349  
484 TTTTGAATGTGCTTCCAGGAGTTCGAATTCAGCGGTGAGAGTGGCAGCTGT 425  
350 CATATGAAGCCACACCAAGCCAGACAGG -TCCTGTTGTCAATCTGTGATGATGAA 408  
424 CATATGAAGCCACACCAAGCCAGACAGG -TCCTGTTGTCAATCTGTGATGATGAA 365  
409 GTTTGAGGGGAAACAAACAGGGGCTTCAACCAAGTCAATCTTCTGACGCCAGGCTTC 468  
364 GTTTGAGGGGAAACAAACAGGGGCTTCAACCAAGTCAATCTTCTGACGCCAGGCTTC 305  
469 ACCCAGAACACAGTGTGGAAGATGCAAGTACTGCTTCCGCTTCCAGAGTGGGCGAG 528  
304 ACCCAGAACACAGTGTGGAAGATGCAAGTACTGCTTCCGCTTCCAGAGTGGGCGAG 245  
529 CTAGTGGGGGTGGCAGAGTCTCTTTCCTTCAATTCAGCCCTAGCTGTAGAAATGCA 588

DB	CTAATGGGGCGGGSAGAGTGTGCTTTGGGTCATTTCAGCCCTTACAGCTCTGTAGAGAAACGA	185
OY	244	
OY	589	AACTCGACTCTCAAGATGTGAGACACACAAGTTCAATTTCTGTGTTCGGAGACACTG
Db	184	AACCTCGACTCTCAAGATGTGAGACACACAAGTTCAATTTCTGTGTTCGGAGACACTG
OY	649	CAGACTCCACTGTGTCGGAGTGTGAACCTTTTGTGTCCTCAAGTTTCAGAGTCCCTT
Db	124	CAGACTCCACTGTGTCGGAGTGTGAACCTTTTGTGTCCTCAAGTTTCAGAGTCCCTT
OY	709	CCCTGAATATATACCTGTTTGTCAATAGTTTCCCTTTTCAAGTAGTAACCTTTTCAATTTT
Db	64	CCCTGAATATATACCTGTTTGTCAATAGTTTCCCTTTTCAAGTAGTAACCTTTTCAATTTT
OY	769	CTAC 772
Db	4	CGAC 1
RESULT 3		
LOCUS	BM975482/c	
DEFINITION	BM975482	685 bp mRNA linear EST 21-MAR-2002
ACCESSION	U1-CF-EN1-acv-k-12-0-UI-s1	U1-CF-EN1 Homo sapiens cDNA clone
VERSION	BM975482	
KEYWORDS	BM975482.1	GI:19593073
SOURCE	EST.	
ORGANISM	human.	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 685)	
JOURNAL	Bonaldo,M.F., Lennon,G. and Soares,M.B.	
MEDLINE	Normalization and subtraction: two approaches to facilitate gene	
COMMENT	discovery	
	Genome Res. 6 (9), 791-806 (1996)	
	9704447	
	Contact: McCray, PB	
	McCray lab	
	University of Iowa	
	2024 University of Iowa Med Labs, Iowa City, IA 52242, USA	
	Tel: 319 356 4866	
	Fax: 319 356 7171	
	Email: paul-mccray@uiowa.edu	
	Tissue Procurement: Dr. M. J. Welsh, University of Iowa	
	cDNA library preparation: Dr. M. Bento Soares, University of Iowa	
	cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa	
	Clone Distribution: Researchers may obtain clones from Research	
	Genetics (www.resgen.com).	
	The following repetitive elements were found in this cDNA	
	sequence: 1-27, >AT-rich#low-complexity (matched complement)	
	Seq primer: M13 FORWARD	
	POLY-A-yes:	
FEATURES		
source	Location/Qualifiers	
	1..685	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="U1-CF-EN1-acv-k-12-0-UI"	
	/clone_11b="U1-CF-EN1"	
	/tissue_type="Primary Lung Cystic Fibrosis Epithelial	
	Cells"	
	/dev_stage="Adult"	
	/lab_host="DH10B (Life Technologies) (TL phase resistant)"	
	/note="Organ: Lung; Vector: pRTT3-Pac (Pharmacia) with a	
	modified polylinker; Site_1: Ecor I; Site_2: Not I;	
	U1-CF-EN1 is a normalized cDNA library containing the	
	following tissue(s): Primary Lung Cystic Fibrosis	
	Epithelial Cells. The library was constructed according to	
	Bonaldo, Lennon and Soares, Genome Research, 6:791-806,	
	1996. First strand cDNA synthesis was primed with an	
	oligo-dt primer containing a Not I site. Double stranded	
	cDNA was ligated to an Ecor I adaptor, digested with Not	

I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dfr)18 tail. The sequence tag for this library is CTGCTCAGCT. TAG, LIB-UI-CF-EN1 TAG, TISSUE-Human Lung Epithelial Cell Lines untreated LPS 6hr to LPS 24h

BASE COUNT 184 a 164 c 163 g 174 t  
ORIGIN  
TAG-SEQ-CTGCTCAGCT"

Query Match 77.7%; Score 645; DB 14; Length 685;  
Best Local Similarity 99.7%; Pred. No.3.4e-168;  
Matches 667; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Oy	104	CCAGAGATGCGATCTGTTGATTTTCAGACCTATGTGATCAGGCTCGAGACCTGCTAG	163
Dd	685	CCAGAGATGCGATCTGTTGATTTTCAGACCTATGTGATCAGGCTCGAGACCTGCTAG	626
Oy	164	GAGTTTGTCAATGCTCTACTACACCAACCATGATTAAGCGCGGCGTTTCTGTCCGCGCTG	223
Dd	625	GAGTTTGTCAATGCTCTACTACACCAACCATGATTAAGCGCGGCGTTTCTGTCCGCGCTG	566
Oy	224	TACATGGGACAGCCACCCCTGGTCTGATGGAATGGCAATGCTGTTTCAGGCAAGATCTTGG	283
Dd	565	TACATGGGACAGCCACCCCTGGTCTGATGGAATGGCAATGCTGTTTCAGGCAAGATCTTGG	506
Oy	284	AGTAGTTTTTGAAGATGTTGCTTCCAGGAGTTCCAAATCAGCGTGTAGACTGCCAG	343
Dd	505	AGTAGTTTTTGAAGATGTTT-CCTTCCAGGAGTTCCAAATCAGCGTGTAGACTGCCAG	447
Oy	344	CCTGTCTATGATGAAGCCACACACCAAGCCAGACACAGGTCCTGTTGTCATCTGTGATCA	403
Dd	446	CCTGTCTATGATGAAGCCACACCAAGCCAGACACAGGTCCTGTTGTCATCTGTGATCA	387
Oy	404	GTGAAGTTTGAAGGGGAACCAACCAAGGGGACTTCAACCAAGAACTTCTATCTGACGCCAG	463
Dd	386	GTGAAGTTTGAAGGGGAACCAACCAAGGGGACTTCAACCAAGAACTTCTATCTGACGCCAG	327
Oy	464	GCCTCACCCACACACACAGTGTGGAAGATGCCAAGTACGCTGTCGCTTCCAGAGACTGG	523
Dd	326	GCCTCACCCACACACACAGTGTGGAAGATGCCAAGTACGCTGTCGCTTCCAGAGACTGG	267
Oy	524	GCCAGCTAGTGGGGTGGCAGAGGCTCTTTGCTTCATTACAGCCCTAGCTCTGTAGAGAA	583
Dd	266	GCCAGCTAGTGGGGTGGCAGAGGCTCTTTGCTTCATTACAGCCCTAGCTCTGTAGAGAA	207
Oy	584	ATGCAAACTCTGACTCTCAAGAGATGTGAAGAACACAACTTATCTTGTGTGGGAGA	643
Dd	206	ATGCAAACTCTGACTCTCAAGAGATGTGAAGAACACAACTTATCTTGTGTGGGAGA	147
Oy	644	CACATGCAAGCTCCACTGTGCCGAGGTAACTCTTTTGTGCTCAAGTCTAGAGATC	703
Dd	146	CACATGCAAGCTCCACTGTGCCGAGGTAACTCTTTTGTGCTCAAGTCTAGAGATC	87
Oy	704	CCTTCCGAAATATATATCTGTTTGTCTATGTTCCCTTTC-AAAGTAGTAACTTCT	762
Dd	86	CCTTCCGAAATATATATCTGTTTGTCTATGTTCCCTTTC-AAAGTAGTAACTTCT	27
Oy	763	ATTTTCTCA 771	
Dd	26	ATTTTCTCA 18	

RESULT 4  
BM543865 1110 bp mRNA linear EST 20-FEB-2002  
LOCUS AGENCOURT 6492315 NIH\_MGC\_125 Homo sapiens cDNA clone IMAGE:558920702  
DEFINITION 5', mRNA sequence.  
ACCESSION BM543865  
VERSION BM543865.1 GI:18774629  
KEYWORDS EST.

Source	Organism	Human
REFERENCE	1 (bases 1 to 1110)	
AUTHORS	NIH-MGC	http://mgc.ncl.nih.gov/
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bbs-remail.nih.gov Tissue Procurement: Invitrogen cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Plate: LLM11361, row: 1 column: 16 High quality sequence stop: 563.	
FEATURES	Location/Qualifiers	
Source	1..1110	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:5589207"	
	/clone_11b="NIH_MGC_125"	
	/lab_host="DH10B"	
	/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6; Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."	
BASE COUNT	249 a 310 c 282 g 259 t 10 others	
ORIGIN		
Query Match	76.2%; Score 632.8; DB 13; Length 1110;	
Best Local Similarity	90.9%; Pred. No. 9.7e-165;	
Matches 719; Conservative	0; Mismatches 63; Indels 9; Gaps 4;	
QY	41 CCGTGCACCCCTGGTTCGCCAAGGACAGAGAAATCCCTGGTGGAGCCCTCTCCATA	100
DB	117 CCGCGCACCCCTGGTTCGCCAAGGACAGAGAAATCCCTGGTGGAGCCCTCTCCATA	176
QY	101 GAACAGAGATGGCATCTGTGGATTTCAGACCTATGTGGATCAGCCTCGACAGCTGCT	160
DB	177 GAACAGAGATGGCATCTGTGGATTTCAGACCTATGTGGATCAGCCTCGACAGCTGCT	236
QY	161 GAGAGATTGTCAATGTCTACTACACCAACATGATTAAGCGGCGGCGTTGCTGTCGCCG	220
DB	237 GAGAGATTGTCAATGTCTACTACACCAACATGATTAAGCGGCGGCGTTGCTGTCGCCG	296
QY	221 CTGTACATGGGACACGCCACCCCTGGTCTGGATGGCATGCTGTTTCAGACACAAATTC	280
DB	297 CTGTACATGGGACACGCCACCCCTGGTCTGGATGGCATGCTGTTTCAGACACAAATTC	356
QY	281 TTGATGAGTTTGTGAATGATGTCCTTCACAGCGAGTTCCAAATCAGGTGGTAGATGCG	340
DB	357 TTGATGAGTTTGTGAATGATGTCCTTCACAGCGAGTTCCAAATCAGGTGGTAGATGCG	416
QY	341 CAGCGCTGTTCATGATGAAGCCACACCAAGCAGACACGCGTCTGTTGTCTATCTGTGGA	400
DB	417 CAGCGCTGTTCATGATGAAGCCACACCAAGCAGACACGCGTCTGTTGTCTATCTGTGGA	476
QY	401 TCAGTGAAGTTTGAAGGGGACACAAACAGGGACTTCAACCCAGACTTATCTTGACCGC	460
DB	477 TCAGTGAAGTTTGAAGGGGACACAAACAGGGACTTCAACCCAGACTTATCTTGACCGC	536
QY	461 CAGGCGTACCCAGACACACAGTGTGGAAGATCGCAATGATGCTTCGCGTTCAGAGC	520
DB	537 CAGGCGTACCCAGACACACAGTGTGGAAGATCGCAATGATGCTTCGCGTTCAGAGC	596

OY		521	TGGGCGACAGTATGTTGGGGGTGGCACAAGGTCCTTGCTTCATATGACCCCTAGCTCTGAGA	580
Db		597	TGGGCACAGTATGTTGGGGGTGGCACAAGGTCCTTGCTTCATATGACCCCTAGCTCTGAGA	654
OY		581	GAAATGCAACCTCGACTCTCAAGAGTGTGAGAACACAAGTTCATTCTGTGTTCGGG	640
Db		655	GAAATGCAACCTCGACTCTCAAGAGTGTGAGAACACAAGTTCATTCTGTGTTCGGG	714
OY		641	AGACACTGCAGACTCTCACCTGTGCCGAGGTTGAACCTTTTTTGTGCTCAAGTTCTAGGA	700
Db		715	GAAACACTGCAGACTCTCACCTGTGCCGAGGTTGAACCTTTTTTGTGCTCAAGGCTTAGGA	774
OY		701	GTCCCTTCCGATATATAAATCTT--GTTTGTGATAGTATCTCTTTCAAGAGTAAATCTT	758
Db		775	GTCCCTTCCGATATATAAATCTT--GTTTGTGATAGTATCTCTTTCAAGAGTAAATCTT	834
OY		759	TT---CTATTTTCTACTTGTGCCAGTGTAGAGACTC--TGATCTCGAAATTCCTGCACAATA	813
Db		835	TTTCATTTTCTTCCAACTGCGCCAGTAATAAATCCTGTGATCTCTGAAATATCTGCACAAT	894
OY		814	ATTTAATATATA 824	
Db		895	AATTTAATATA 905	
RESULT 5				
LOCUS	BPI95575/c			
DEFINITION	BPI95575	720 bp	mRNA	linear EST 03-NOV-2000
	7n55b1.x1 NCI-CGAP.Ov18 Homo sapiens cDNA clone IMAGE:3571173 3'			
	similar to TR:Q9UKK6 Q9UKK6 NTF2-RELATED EXPORT PROTEIN NPTL.1 ;			
	mRNA sequence.			
ACCESSION	BPI95575			
VERSION	BPI95575.1	GI:11082605		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
AUTHORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
TITLE	Tumor Gene Index			
JOURNAL	Unpublished (1997)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center			
	Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNLB, send email to: info@image.lnl.gov			
FEATURES	High quality sequence stop: 466.			
source	Location/Qualifiers			
	1..720			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone_image="IMAGE:3571173"			
	/clone_id="NCI-CGAP.Ov18"			
	/tissue_type="fibrocyteoma"			
	/lab_host="DH10B (phage-resistant)"			
	/note="Organ: ovary; Vector: pTZ19-Pac (Pharmacia) with a modified polylinker; Site.1: Not.I; Site.2: Eco.RI; 1st strand cDNA was primed with a Not.I -Oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCCGACGCAATTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco.RI adaptors (Pharmacia), digested with Not.I and cloned into the Not.I and Eco.RI sites of the modified pTZ19 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."			
BASE COUNT	205 a 170 c 169 g 175 t 1 others			





Db 31 TTTCAAGTAGTAACCTTTCTATTCTTCT 1

RESULT 7  
BM789812 649 bp mRNA linear EST 05-MAR-2002  
LOCUS K-EST0069490 S22SNU16 Homo sapiens cDNA clone S22SNU16-1-G02 5'

DEFINITION  
mRNA sequence.  
BM789812  
BM789812.1 GI:19138044

ACCESSION  
BM789812

VERSION  
EST.

KEYWORDS  
human.

SOURCE  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,  
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and:  
Kim, Y.S.  
21C Frontier Korean EST Project 2001  
Unpublished (2002)  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: Yongsung@mail.kribb.re.kr  
Plate: 1 row: 6 column: 02  
High quality sequence stop: 649.  
Location/Qualifiers

FEATURES  
source  
1. 649  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S22SNU16-1-G02"  
/clone\_11b="S22SNU16"  
/sex="F"  
/tissue\_type="Ascites"  
/cell\_type="Lymphoblast-like"  
/cell\_line="SNU-16"  
/lab\_host="DH10B"  
/note="Organ: Stomach; Vector: pT73-Pac; Site: 1; EcoRI;  
Site-2: NotI; The S22SNU16 library was contributed by the  
Soares Laboratory and it was constructed as described by  
Bonald, M.F., Lennon, G. and Soares, M.B. (1996), Genome  
Research 6(9): 791-806. RNA was prepared from harvested  
cells of SNU-16 culture. SNU-16 cell was obtained from  
Korean Cell Line Bank (KCLB). SNU-16 was established from  
ascitic fluids of Korean patients by Park J.G. et al.  
(1990), Cancer Res 50: 2773-2780."

BASE COUNT  
156 a 171 c 172 g 150 t

ORIGIN

Query Match 75.1%; Score 623; DB 14; Length 649;  
Best Local Similarity 100.0%; Pred. No. 4,4e-162;  
Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

49 CCCCCTGTTCCCAAGCAGAGGAATACCTGCTGAGCCCTCTTCATAGAACCGA 108  
DB 27 CCCCCTGTTCCCAAGCAGAGGAATACCTGCTGAGCCCTCTTCATAGAACCGA 86  
109 GATGCACTGTGATTTCAAGACCTATGTGATCAGGCTTCAGAGCTGTGAGAGTT 168  
DB 87 GATGCACTGTGATTTCAAGACCTATGTGATCAGGCTTCAGAGCTGTGAGAGTT 146  
169 TGTCAATGTCTACTACACCATGATAGAGGGGGGGCTTGTCTCCCGCTTACAT 228  
DB 147 TGTCAATGTCTACTACACCATGATAGAGGGGGGGCTTGTCTCCCGCTTACAT 206  
229 GGGCAGACGACCCGTGCTGGAATGCAATCTTTTCAGACAAAGATCCTTGAGTGA 288  
DB 207 GGGCAGACGACCCGTGCTGGAATGCAATCTTTTCAGACAAAGATCCTTGAGTGA 266

QY 289 GTTTTGAATGTCCTCCAGCAGAGTTCGAATACGCTGTAGACTGCCAGCTGT 348  
DB 267 GTTTTGAATGTCCTCCAGCAGAGTTCGAATACGCTGTAGACTGCCAGCTGT 326  
QY 349 TCATGATGAGCCACACCAAGCCAGACGCTGCTGTGTGATCTGTGATGATGA 408  
DB 327 TCATGATGAGCCACACCAAGCCAGACGCTGCTGTGTGATCTGTGATGATGA 386  
QY 409 GTTTGAGGGGACAAACACCGGACTTCAACGAACTTCATCCAGCCAGCCCTC 468  
DB 387 GTTTGAGGGGACAAACACCGGACTTCAACGAACTTCATCCAGCCAGCCCTC 446  
QY 469 ACCCAGACACAGTGTGAGATCCAGTGAAGTCTCCGCTTCAGAGCTGGCCAG 528  
DB 447 ACCCAGACACAGTGTGAGATCCAGTGAAGTCTCCGCTTCAGAGCTGGCCAG 506  
QY 529 CTAGTGGGGGTGGCAGAGGCTCTTTGCTTCATTCAGCCCTAGCTGTGAGAAATGA 588  
DB 507 CTAGTGGGGGTGGCAGAGGCTCTTTGCTTCATTCAGCCCTAGCTGTGAGAAATGA 566  
QY 589 AACCTGACTCTCAAGATGTGAGAGACACAAATCTTCTGTTGGGGAGACACTG 648  
DB 567 AACCTGACTCTCAAGATGTGAGAGACACAAATCTTCTGTTGGGGAGACACTG 626  
QY 649 CAGACTCCACTGTGCGGAGTTG 671  
DB 627 CAGACTCCACTGTGCGGAGTTG 649

RESULT 8  
B0581596/c 644 bp mRNA linear EST 20-JUN-2002  
LOCUS B0581596  
DEFINITION  
1110905.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6029745  
3' similar to IR:Q9UKK6 Q9UKK6 NIF2-RELATED EXPORT PROTEIN NMT1.  
mRNA sequence.  
B0581596  
B0581596.1 GI:21494492  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 644)  
Lemstra, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
Lemstra, L., Scaer, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blisat, A.,  
Schmitt, A., Theising, B., Rittler, E., Ronko, I., Bennett, J., Gardnas  
, M., Gibbons, M., McCann, R., Cole, R., Tsagarelis, R., Williams, T.,  
Jackson, Y. and Bowers, Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@hms.harvard.edu  
Library was constructed by Dr. J. Ferrer in vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
University Genome Sequencing Center For information on obtaining a  
Seq primer: -400P from Gibco  
High quality sequence stop: 488.  
Location/Qualifiers

FEATURES  
source  
1. 644  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_11b="Human insulinoma"  
/tissue\_type="Insulinoma"

```

/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pBluescript SK-; Site_1:
XhoI; Site_2: EcoRI; Constructed with lambda ZapII system
(Stratagene) by Dr. J. Ferrer. In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington
University protocol
(http://genome.wustl.edu/est/lambda-protocol.shtml) .
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library. "

```

BASE COUNT	179 a	142 c	149 g.	174 t
ORIGIN				

Query Match	73.9%	Score 613	DB 14	Length 644
Best Local Similarity	100.0%	Prod. No. 2.6e-159		
Matches 613	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	218	CGCCTGTCATAGGCGACACAGCCACCCCTGGTCTGGGAATGGCATGCTGTTTCAGACACAGAA	277	
DB	644	CGCCTGTCATAGGCGACACAGCCACCCCTGGTCTGGGAATGGCAATGCTGTTTCAGACACAGAA	585	
QY	278	TCCTTGATGAGTATTTTGGAAATGTTGGCTTCCAGCGAGTCCCAATCAGCGGTGTAGAC	337	
DB	564	TCCTTGATGAGTATTTTGGAAATGTTGGCTTCCAGCGAGTCCCAATCAGCGGTGTAGAC	525	
QY	338	TGCCAGCCTGTCATGATGAAGCCACACCAAGCCAGACCCAGCGTCTTGTGTCACTCTGT	397	
DB	524	TGCCAGCCTGTCATGATGAAGCCACACCAAGCCAGACCCAGCGTCTTGTGTCACTCTGT	465	
QY	398	GGATAGTGAAGTTGAGGGGAAACAAGGGGACTCAACAGAACTTATCCATGAC	457	
DB	464	GGATAGTGAAGTTGAGGGGAAACAAGGGGACTCAACAGAACTTATCCATGAC	405	
QY	458	GCCCAAGCCTCACCAGCAACACAGTGTGAAGATCGCAAGTACTGCTCCGCTTCAG	517	
DB	404	GCCCAAGCCTCACCAGCAACACAGTGTGAAGATCGCAAGTACTGCTCCGCTTCAG	345	
QY	518	GACTGGGCGCACTACTGGGGGGGAGAGAGTCTCTTGCTTCATCAGCCCTAGCTGCT	577	
DB	344	GACTGGGCGCACTACTGGGGGGGAGAGAGTCTCTTGCTTCATCAGCCCTAGCTGCT	285	
QY	578	AGAGAAATGACAAACCTCGACACTCTCAAGATGTGAGAGAACACAAAGTTCAATTTCTGTGTG	637	
DB	284	AGAGAAATGACAAACCTCGACACTCTCAAGATGTGAGAGAACACAAAGTTCAATTTCTGTGTG	225	
QY	638	CGGAGACACTGACGACTCCACTGTGCGCAGAGTTGAACCTCTTTTGTGTGCTCAAGTCTTA	697	
DB	224	CGGAGACACTGACGACTCCACTGTGCGCAGAGTTGAACCTCTTTTGTGTGCTCAAGTCTTA	165	
QY	698	GGAGTCCCTTCTGTAATATATACGTGTTGTGCATAGTTCTCTTTTCAAGTAGTAACCT	757	
DB	164	GGAGTCCCTTCTGTAATATATACGTGTTGTGCATAGTTCTCTTTTCAAGTAGTAACCT	105	
QY	758	TTTCTATTTTCTACTTGCCAGTAGAGAGACTGTGATCTGAAATTTCTGACAAATTAATTT	817	
DB	104	TTTCTATTTTCTACTTGCCAGTAGAGACTGTGATCTGAAATTTCTGACAAATTAATTT	45	
QY	818	AATAATACACATG 830		
DB	44	AATAATACACATG 32		

RESULT	9
BGI09830	
LOCUS	BGI09830
DEFINITION	BGI09830 602208099f1 NTH_MGC_86 Homo sapiens CDNA clone IMAGE:4368311 5'
ACCESSION	mRNA sequence.
VERSION	BGI09830
KEYWORDS	BGI09830.1 GI:12603336
SOURCE	Est.
	human.

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Homo sapiens					
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:					
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.					
1 (bases 1 to 906)					
NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .					
National Institutes of Health, Mammalian Gene Collection (MGC)					
Unpublished (1999)					
Contact: Robert Strausberg, Ph.D.					
Email: <a href="mailto:cgapbs@email.nih.gov">cgapbs@email.nih.gov</a>					
Tissue Procurement: ATCC					
CDNA Library Preparation: Life Technologies, Inc.					
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)					
DNA Sequencing by: Incyte Genomics, Inc.					
Clone distribution: MGC clone distribution information can be					
found through the I.M.A.G.E. Consortium/LNL at:					
<a href="http://image.llnl.gov">http://image.llnl.gov</a>					
Plate: LHAM10022 row: b column: 24					
High quality sequence stop: 705.					
Location/Qualifiers					
1..906					

```

/clone="IMAGE:4368311"
/clone_1b="NIH_MGC_86"
/tissue_type="osteosarcoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: bone; Vector: pCMV-SPORT6; Site.1: NotI.
Site.2: SalI; Cloned unidirectionally; oligo-df primed.
Average insert size 1.533 kb. Library enriched for
full-length clones and constructed by Life Technologies
Note: this is a NIH_MGC Library."

```

Query Match	Similarity	93.4%	Score	609	DB	12	Length	906
Best Local	Similarity	93.8%	Pred. No.	3,7e-158				
Matches	678	Conservative	0	Mismatches	40	Indels	5	Gaps
QY	41	CCCTGACGCCCTGGTGTCTTCCCAAGCAGCAGAGAATAACCTGCTGGAGCCCTCTTCCATA	100					
DB	118	CCGCCACGCCCTCTGGTGTCTCCCAAGCAGAGGAAATACCTGGTGAAGCCCTCTTCCATA	177					
QY	101	GACCCAGAGATGGCATCTGTGATTTCAAGACCTTATGTGGATCAGAGCCCTGCAGACTGCT	160					
DB	178	GACCCAGAGATGGCATCTGTGATTTCAAGACTTATGTGGATCAGAGCCCTGCAGACTGCT	237					
QY	161	GAGAGATTGTCAAGTCTACTACACCAACCATGGATTAAGCGGCGGCTTGTCTGCCGC	220					
DB	238	GAGAGATTGTCAAGTCTACTACACCAACCATGGATTAAGCGGCGGCTTGTCTGCCGC	297					
QY	221	CTGTACATGGGCACAGCCACCCTGCTCTGGAAATGCAATGCTGTTTCAGGACAAATCC	280					
DB	298	CTGTACATGGGCACAGCCACCCTGCTCTGGAAATGCAATGCTGTTTCAGGACAAATCC	357					
QY	281	TTCAGTGAAGTTTGTGAATGTTGGCTTCCACAGGATTCCAATAGCGTGTAGACTGC	340					
DB	358	TTCAGTGAAGTTTGTGAATGTTGGCTTCCACAGGATTCCAATAGCGTGTAGACTGC	417					
QY	341	CAGCGTGTTCATGATGAAGCCACACCAAGCCAGACGACGCTCTTGTTCATCTGTGGA	400					
DB	418	CAGCGTGTTCATGATGAAGCCACACCAAGCCAGACGACGCTCTTGTTCATCTGTGGA	477					
QY	401	TCAGTGAAGTTTGTGAGGGACAACAACGGGACTTCAACCAATCTCATCTGTGACCGC	460					
DB	478	TCAGTGAAGTTTGTGAGGGACAACAACGGGACTTCAACCAATCTCATCTGTGACCGC	537					
QY	461	CAGGCGCTCACGAGAACACAGTGGGAAGATCCGAAGAGTCTGCTCCGCTTCAGAGAC	520					
DB	538	CAGGCGCTCACGAGAACACAGTGGGAAGATCCGAAGTCTGCTCCGCTTCAGAGAC	597					
QY	521	TGGCGCAGCTACTGGGGGTGGCAGAGGCTCTCTTTGCTTCATTACAGCCCTACCTGTGAG	579					
DB	598	TGGCGCAGCTACTGGGGGTGGCAGAGGCTCTCTTTGCTTCATTACAGCCCTACCTGTGAG	655					

Query Match	Best Local Similarity	Score	DB	Length
72.5%	99.2%	602	10	639
Matches	605	Conservative	0	Mismatches
5	Indels	0	Gaps	0

Query	Match	Score	DB	Length
140	GATCAGCGCGCAAGCTGCTGAGAGCTTTGTCAATGCTACTACACACATGATGATG	199		
1	GATAGAGCCCTGCAAGCTGCTGAGAGATTTGTTCATGCTTACTACACACATGATGATG	60		
200	CGCGCGCGCTTGTCTGTCGCCGCTGTACTACAGGCGCACACCCCTGCTCGAATGCGCAT	259		

[illegible]



```

Db      420 TCATCACCCTAGCTCTGTAGAGAAATGCAACCTCAGCTCTCAGAGATGTGAGAACCA 479
Oy      618 CAAGTTCATTTCTGTGTGGGAGACACATCGACACTGCTGCGAGGTGTAACCTCT 677
Db      480 CAATTCATTTCTGTGTGGGAGACACATCGACACTGCTGCGAGGTGTAACCTCT 538
Oy      678 TTTTGTGCTCAGAGTCTAGAGTCCCTTCTCTGATATATATCTGTTTGTCAATGTTT 737
Db      539 TTTTGTGTTTCAAGTTCTAGAGTCCCTTCTCTGATATATATCTGTTTGTCAATGTTT 597
Oy      738 CCTTTCAAGTACTAACTTTCTATTTT 767
Db      598 CCTTTCAAGTACTAACTTTCTATTTT 627

RESULT 13
LOCUS   BE794429
DEFINITION 601589445F1 NIH_MGC_7 Homo sapiens CDNA clone IMAGE:3943762 5'
ACCESSION BE794429
VERSION   BE794429.1 GI:10215627
KEYWORDS EST.
SOURCE    human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 882)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgaabs-remail.nih.gov
          Tissue Procurement: DCTD/DRP
          CDNA Library Preparation: Ling Hong/Rubin Laboratory
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov
          Plate: LICM799 row: 1 column: 11
          High quality sequence stop: 756.
FEATURES
Source    location/Qualifiers
1..882
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3943762"
/clone.lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOT7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dr priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACAG(G). Size selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 190 a 246 c 237 g 209 t
ORIGIN
Query Match 71.3%; Score 591.8; DB 12; Length 882;
Best Local Similarity 97.0%; Pred. No. 2.2e-153;
Matches 668; Conservative 0; Mismatches 12; Indels 9; Gaps 6;
Oy      41 CCTTCGACGCTGTGTGTTCCCAAA-GCAGAGAAATACCTGCTGAGAGCTCTTCAT 99
Db      137 CCGCCGACCCCTGTGTTCCCAAGGAGAGAAATACCTGCTGAGAGCTCTTCAT 196
Oy      100 AGAACGAGATGATGATGATTTCAAGACCTATGATGATGAGGCTCGAGAGCTGC 159
Db      197 AGAACGAGATGATGATGATTTCAAGACCTATGATGATGAGGCTCGAGAGCTGC 256

```

```

Oy      160 TGAGAGTTTGTCAATGCTACTACACCATATGATAAGGGCGGCTTGTCTCCCG 219
Db      257 TGAGAGTTTGTCAATGCTACTACACCATATGATAAGGGCGGCTTGTCTCCCG 316
Oy      220 CCTGTACATGGGCAAGCACCCTGTGTGTAATGCAATGCTGTTTGTGAGCAAGATC 279
Db      317 CCTGTACATGGGCAAGCACCCTGTGTGTAATGCAATGCTGTTTGTGAGCAAGATC 376
Oy      280 CTTAGAGAGTTTGTGAAATGTTGCTTCCAGCAGTTTCAAAATCAGCGTGTAGACTG 339
Db      377 CTTAGAGAGTTTGTGAAATGTTGCTTCCAGCAGTTTCAAAATCAGCGTGTAGACTG 436
Oy      340 CCAGCTGTTTATGATGATGAGCCACACCAAGCAGCAGGCTCTTGTGTGATCTGTGG 399
Db      437 CCAGCTGTTTATGATGATGAGCCACACCAAGCAGCAGGCTCTTGTGTGATCTGTGG 496
Oy      400 ATCAGTGAAGTTTGTGAGGGAACAACAACGAGCTTCAACGAGATTCATCTGACCGC 459
Db      497 ATCAGTGAAGTTTGTGAGGGAACAACAACGAGCTTCAACGAGATTCATCTGACCGC 556
Oy      460 CCAAGCTTCAACCCAGCAACACAGTGTGAAGATGCAAGTGTCTTCCGCTTCCAGGA 519
Db      557 CCAAGCTTCAACCCAGCAACACAGTGTGAAGATGCAAGTGTCTTCCGCTTCCAGGA 616
Oy      520 CTGGGCCAGCTAGTGGGGG-TGGCAGAGTCTCTTCTTCTTCAATCAGCCCTAGCTGTA 578
Db      617 CTGGGCCAGCTAGTGGGGGTTGGCAGAGTCTCTTCTTCAATCAG-CTAGCTGTGTA 675
Oy      579 GAGAAATGCAAACTGACCTCTCAA-GAATGTAGGAACAACAAGTTATTTCTTTG 637
Db      676 GAGAAATGCAAACTGACCTCTCAAAGGATGTAGGAACAACAAGTTATTTCTG 734
Oy      638 CGAGAGACTGACAGACTCCAGCTGCGAGGTGCAACTCTTTTGTGTCAGAGTCTTA 697
Db      735 CGAGAGACTGACAGACTCCAGCTGCGAGGTGCAACTCTTTTGTGTCAGAGTCTTA 790
Oy      698 GAGAGCTCTTCTCTGAAATATATCTGTT 726
Db      791 GAGAGCTCTTCTCTGAAATATATCTGTT 819

RESULT 14
LOCUS   BE463515/c
DEFINITION hw24e08.x1 NCI_CGAP_Kid1 Homo sapiens CDNA clone IMAGE:3183878 3'
ACCESSION BE463515
VERSION   BE463515.1 GI:9509290
KEYWORDS EST.
SOURCE    human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 606)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgaabs-remail.nih.gov
          Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
          Emmerit-Buck, M.D., Ph.D.
          CDNA Library Preparation: M. Bento Soares, Ph.D.
          CDNA Library Arrayed by: Greg Lennon, Ph.D.
          DNA Sequencing by: Washington University Genome Sequencing Center
          found through the I.M.A.G.E. Consortium/LNL, send email to:
          InfoImage.lnl.gov
          Seq primer: -400p from Gibco
          High quality sequence stop: 453.
FEATURES
Source    location/Qualifiers
1..606

```

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3183878"  
/clone\_lib="NCI-CGAP\_Kid11"  
/note="Organ: Kidney; Vector: p773p-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; plasmid DNA from the normalized library NCI-CGAP\_Kid3 was prepared, and 88 circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 132376-132391, 1456007-1456775, and 1500552-1502853). Subtraction by Bento Soares and M. Fatima Bonalido."

BASE COUNT 171 a 144 c 150 g 141 t

Query Match 71.2%; Score 590.6; DB 10; Length 606;  
Best Local Similarity 98.5%; Pred. No. 4.3e-153;  
Matches 596; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

167 TTGTGATGCTCTACTACACCATGATTAAGCGGCGCTTGTCTCCCGCTGTAC 226  
|||||  
606 TTGTGATGCTCTACTACACCATGATTAAGCGGCGCTTGTCTCCCGCTGTAC 547  
|||||

227 ATGGGACAGCCACCTGGTCTGGAATGCAATGCTTTTCAGACAAGAACTCTGAGT 286  
|||||  
546 ATGGGACAGCCACCTGGTCTGGAATGCAATGCTTTTCAGACAAGAACTCTGAGT 487  
|||||

287 GAGTTTTTGAAGTGGCTTCCAGAGATTCGAATCAGCGTGTAGATGCGAGCT 346  
|||||  
486 GAGTTTTTGAAGTGGCTTCCAGAGATTCGAATCAGCGTGTAGATGCGAGCT 427  
|||||

347 GTTCATGATGAAGCACACCAAGCAGCGGCTTGTTCATCTGTGATCACTG 406  
|||||  
426 GTTCATGATGAAGCACACCAAGCAGCGGCTTGTTCATCTGTGATCACTG 367  
|||||

407 AAGTTTGGGGGAACAACAAGGGAATTCACAGAACTTCATCTGACCGCCAGGCC 466  
|||||  
366 AAGTTTGGGGGAACAACAAGGGAATTCACAGAACTTCATCTGACCGCCAGGCC 307  
|||||

467 TCACCCAGCAACAGATGGAAGATGCAAGTACCTGCTCCAGAGCTGGGCC 526  
|||||  
306 TCACCCAGCAACAGATGGAAGATGCAAGTACCTGCTCCAGAGCTGGGCC 247  
|||||

527 AGCTAGTGGGGGTGGCAGAGTCTCTTGTTCATTCAGCCCTAGCTGTAGAGAAATG 586  
|||||  
246 AGCTAGTGGGGGTGGCAGAGTCTCTTGTTCATTCAGCCCTAGCTGTAGAGAAATG 187  
|||||

587 CAAACCTGCACTCTCAAGATGTGAGAAACAAAGTCAATTTCTGTTGGGAGACAC 646  
|||||  
186 CAAACCTGCACTCTCAAGATGTGAGAAACAAAGTCAATTTCTGTTGGGAGACAC 127  
|||||

647 TGCAGACTCACTGTGCGGAGGTGAACCTTTTGTGTCAGAGTTCTAGAGTCCCT 706  
|||||  
126 TGCAGACTCACTGTGCGGAGGTGAACCTTTTGTGTCAGAGTTCTAGAGTCCCT 67  
|||||

707 TTCCGGAATATATCTTTGTTGTCTCAATGTTTCTTCAAGTAACTTTCTATTT 766  
|||||  
66 TTCCGGAATATATCTTTGTTGTCTCAATGTTTCTTCAAGTAACTTTCTATTT 7  
|||||

767 TTCTTA 771  
|||||  
6 TTCTTA 2  
|||||

RESULT 15  
BE616032 596 bp mRNA linear EST 24-Aug-2000  
LOCUS BE616032 601279506f1 N1H\_MGC\_39 Homo sapiens cDNA clone IMAGE:3611641 5'  
DEFINITION mRNA sequence.  
ACCESSION BE616032

VERSION BE616032.1 GI:9897548  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.lnl.gov  
Plate: LCM269 row: k column: 02  
High quality sequence stop: 589.

FEATURES  
source  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3611641"  
/clone\_lib="N1H\_MGC\_39"  
/tissue\_type="adenoecarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 137 a 150 c 154 g 154 t 1 others

Query Match 71.0%; Score 589.4; DB 10; Length 596;  
Best Local Similarity 99.7%; Pred. No. 9.3e-153;  
Matches 590; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

137 GTGATCAGGCTCGAGAGCTCTGAGAGTGTTCATGTCTATACACCACTGAT 196  
|||||  
1 GTGATCAGGCTCGAGAGCTCTGAGAGTGTTCATGTCTATACACCACTGAT 60  
|||||

197 AAGCGGGCGCTTGTCTGCGCGCTGTACATGGGACAGCCCTGCTGGAATGCG 256  
|||||  
61 AAGCGGGCGCTTGTCTGCGCGCTGTACATGGGACAGCCCTGCTGGAATGCG 120  
|||||

257 AATGCTGTTTCAGCAGCAAGATCTTGAAGTGTGTTTGAATGTTGCTTCCAGCAG 316  
|||||  
121 AATGCTGTTTCAGCAGCAAGATCTTGAAGTGTGTTTGAATGTTGCTTCCAGCAG 180  
|||||

317 TTCCAATCAGCGGTGATGCTCCAGACCTGTTTCATGATGAAGCCACACCAAGCAGAC 376  
|||||  
181 TTCCAATCAGCGGTGATGCTCCAGACCTGTTTCATGATGAAGCCACACCAAGCAGAC 240  
|||||

377 AAGGTCCTGTTGTGATCATCTGTGATGAGAGTTGAGAGGGAACAAACAGGGAGCTTC 436  
|||||  
241 AAGGTCCTGTTGTGATCATCTGTGATGAGAGTTGAGAGGGAACAAACAGGGAGCTTC 300  
|||||

437 AACCAAGACTTATCTGACCGCCCGCCAGGCTCACCAGCAACAGCACTGTGAAGATCGCA 496  
|||||  
301 AACCAAGACTTATCTGACCGCCCGCCAGGCTCACCAGCAACAGCACTGTGAAGATCGCA 360  
|||||

497 AGTACTGCTTCCGCTTCCAGAGCTGGCCAGCTATGTGGGGGTGGAGAGTCTCTTTCG 556  
|||||  
361 AGTACTGCTTCCGCTTCCAGAGCTGGCCAGCTATGTGGGGGTGGAGAGTCTCTTTCG 420  
|||||

557 TTCATTCAGCCCTAGCTGTGAGAGAAAGCAAACTGACTCTCAAGATGTGAGAGAC 616  
|||||

Db 421 |||||  
TTCATTCACCCCTAGCTGTGTAGAGAAATGCAAACTCGACTCTCAAGGATGTGAGAAC 480  
QY 617 ACAACTTCATTTCTGTGTGTGGGAGACACTGCAGACTCCACTGTGCCGAGGTGAATC 676  
Db 481 ACAAGTTTCTCTGTGTGTGGGAGACACTGCAGACTCCACTGTGCCGAGGTGAATC 540  
QY 677 TTTTGTGCTCAGCTCTAGAGACTCCCTTTCTGAAATATATACTTGTG 728  
Db 541 TTTTGTGCTCAGCTCTAGAGACTCCCTTTCTGAAATATATACTTGTG 592

Search completed: June 28, 2003, 03:34:20  
Job time: 1399.27 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:39:51 ; Search time 213.801 Seconds  
(without alignments)

8742.532 Million cell updates/sec

Title: US-09-763-902b-10

Sequence: 1 tgacttgcgcattccacgtg.....ataattaataacacatg 830

Scoring table: IDENTITY\_NOC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*

2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*

3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*

4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*

5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*

6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*

7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*

8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*

9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*

10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*

11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*

12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*

13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*

14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*

15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*

16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*

17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*

18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*

19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*

20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*

21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*

22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*

23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*

24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	830	100.0	830	21 AAA08036 Human protein tran
2	761	91.7	864	22 AAS26215 Human CDNA encodin
3	317	38.2	318	23 AAS58525 Human gene signatu
4	297.4	30.9	359	16 AAT22887 Human CDNA clone (
5	256.2	30.3	716	22 AAH07014 Human protein tran
6	251.6	30.3	1146	21 ABA08041 Human NTP associa
7	251.6	30.3	2581	24 ABA05754 Human CDNA sequenc
8	251.6	30.3	2692	22 AAH17938 Human prostate exp
9	251.6	30.3	2994	23 ABV24535

10	250	30.1	1387	22 AAS26214 Human CDNA encodin
11	223	26.9	247	24 ABL62553 Colon adenocarcino
12	132.6	16.0	133	19 AAX10492 Human biallelic po
13	113	13.6	8918	22 AAS26708 Human genomic DNA
14	113	13.6	8919	22 AAS26709 Human genomic DNA
15	105.2	12.7	1035	23 AAS75617 DNA encoding novel
16	76.4	9.2	805	23 ABL16401 Drosophila melanog
17	60	7.2	60	24 ABL38042 Human spliced tran
18	58.8	7.1	2932	23 ABL16400 Drosophila melanog
19	58.8	7.1	4050	23 ABL09616 Drosophila melanog
20	58.8	7.1	7883	23 ABL08308 Drosophila melanog
21	52.8	6.4	65	24 ABL3496 Mouse spliced tran
22	49.4	6.0	51	22 ABL27777 Human SNP oligonuc
23	39	4.7	50	22 ABL34644 Human SNP oligonuc
24	38	4.6	753	21 AAC39195 Zea mays DNA fragm
25	37.2	4.5	16633	24 ABL70286 Human chemically m
26	37	4.5	4993	24 ABL34103 Human immune syste
27	36.2	4.4	1050	22 AAH41175 Mammalian Cre reco
28	36.2	4.4	6117	24 ABL33024 Human immune syste
29	36	4.3	285	21 ABL00696 Human colon cancer
30	36	4.3	9666	24 ABL70286 Human immune syste
31	36	4.3	9666	24 ABL33085 Human metastasis a
32	36	4.3	9666	24 ABL34531 Human metastasis a
33	35.8	4.3	6113	24 ABL70607 Human chemically m
34	35.8	4.3	15518	24 ABL70607 Human immune syste
35	35.8	4.3	15518	24 ABL34172 Human immune syste
36	35.8	4.3	15518	24 ABL34624 Human metastasis a
37	35.8	4.3	32191	22 ABL37009 Human musculoskele
38	35.4	4.3	506	23 ABLV6423 Human prostate exp
39	35.4	4.3	4415	24 ABL70423 Chemically treated
40	35.4	4.3	4415	24 ABL33906 Human immune syste
41	35.4	4.3	4415	24 ABL34602 Human metastasis a
42	35.2	4.2	2149	22 AAK94875 Human full-length
43	35.2	4.2	116524	19 AAV52850 Human eyal gene co
44	35.2	4.2	151826	21 AAV22281 BAC containing rep
45	35	4.2	3156	24 ABL57129 Bacillus subtilis

# ALIGNMENTS

RESULT 1	AAA08036	standard; CDNA; 830 BP.
ID	AAA08036;	
AC	AAA08036;	
XX		
DT	19-JUN-2000	(first entry)
DE	Human protein transport molecule (PTM) encoding CDNA SEQ ID NO:10.	
XX		
KW	Human: protein transport molecule; PTM: diagnosis; cytosolic;	
KW	antiarthritic; antidiabetic; immunosuppressant; antileukemic;	
KW	dermatologic; antidiabetic; antileukemic; antileukemic; osteoporotic;	
KW	antileukemic; antidiabetic; antileukemic; antileukemic; antileukemic;	
KW	antileukemic; antidiabetic; antileukemic; antileukemic; antileukemic;	
KW	immune disorder; cell proliferative disorder; secretory disorder;	
KW	uticaria; allergy; abnormal vesicle trafficking; asthma;	
KW	autoimmune haemolytic anemia; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	MO200012703-A2.	
XX		
PD	09-MAR-2000.	
XX		
PF	26-AUG-1999; 99MO-US19616.	
XX		
PR	27-AUG-1998; 98US-0098206.	
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;	

PI Gorgone GA, Baughn MR, Patterson C;  
XX  
DR WPI: 2000-256647/22.  
DR P-PSDB; NAI82318.  
XX  
XX New human protein transport-associated polypeptide and polynucleotide  
PT useful for diagnosis, prevention and treatment of cell proliferative  
PT and secretory disorders such as leukemia, cystic fibrosis -  
XX  
XX Claim 9; Page 70; 75pp: English  
PS

CC AAA08033 o AAA08042 encode the human protein transport-associated  
CC molecules (PRAMs) given in AY82317 to AY82324. The PRAMs have  
CC cytosolic, antiarthritic, antisthmatic, immunosuppressant,  
CC antiatherosclerotic, antiallergic, antidiabetic, antilipemic,  
CC antirheumatic, osteopathic, dermatological, antineoplastic, antipsoriatic,  
CC hepatotropic, antitumor, antiinflammatory and antiHIV activities, and  
CC regulate protein transport. PRAM proteins and antagonists are useful for  
CC preventing or treating a disorder associated with decreased or increased  
CC expression or activity of PRAM. PRAM polynucleotides are useful for  
CC diagnosing conditions associated with PRAM, comprising detecting PRAM by  
CC forming a hybridisation complex, preferably after PCR amplifying the  
CC biological sample. Diseases prevented, treated or diagnosed include cell  
CC proliferative disorders such as cancers, immune disorders, secretory  
CC disorders and other conditions associated with abnormal vesicle  
CC trafficking, such as allergies, asthma, urticaria and autoimmune  
CC haemolytic anaemia. Anti-PRAM antibodies may be used as antagonists, as  
CC a targeting or delivery mechanism for bringing pharmaceutical agents  
CC into cells or tissues expressing PRAM and for diagnosis of PRAM-related  
CC disorders. PRAM, its catalytic or immunogenic fragments are useful for  
CC drug screening using libraries of compounds. PRAM polynucleotides are  
CC useful for generating hybridisation probes useful in mapping the  
CC naturally occurring genomic sequences.

SQ Sequence 830 BP; 198 A; 208 C; 195 G; 229 T; 0 other;

Query Match	100.0%	Score 830;	DB 21;	Length 830;
Best Local Similarity	100.0%	Pred. No. 2.9e-246;		
Matches 830; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

[illegible]

Db 481 AGTGTGGAAATCCGAAGTGAAGTCTTCCGCTTCCAGAGACTGGGCGCAGCTAGTGGGGGTG 540

Qy 541 GCAGAGTCTCTTTTCTCTTCATTCACGCCCTAGCTCTGTGTAGAGAAATGCAAACTCGACTGT 600

Db 541 GCAGAGTCTCTTTTCTCTTCATTCACGCCCTAGCTCTGTGTAGAGAAATGCAAACTCGACTGT 600

Qy 601 CAAGATGTGAGGACACACAAAGTTCATTCTGTGTGGCGAGACACTGCGAGATCTACTG 660

Db 601 CAAGATGTGAGGACACACAAAGTTCATTCTGTGTGGCGAGACACTGCGAGATCTACTG 660

Qy 661 TGCCGAGGTGAAGTCTTTTTTGTGTGCTCAGTCTAGAGATGCCCTTTCCTGAATATATA 720

Db 661 TGCCGAGGTGAAGTCTTTTTTGTGTGCTCAGTCTAGAGATGCCCTTTCCTGAATATATA 720

Qy 721 CTGTGTTTGCATAGTTTCGTTTCCATCAAGTAGTAAGTTCATTTTCTACTATGCCAG 780

Db 721 CTGTGTTTGCATAGTTTCGTTTCCATCAAGTAGTAAGTTCATTTTCTACTATGCCAG 780

Qy 781 TAGAGACTCTGATTCGTGAAATATTCGACAAATATATTAATTAATATATGACATG 830

Db 781 TAGAGACTCTGATTCGTGAAATATTCGACAAATATATTAATTAATATATGACATG 830

RESULT 2  
AAS26215  
ID: AAS26215 standard; cDNA; 864 BP

AC AAS26215;

DT 07-NOV-2001 (first entry)

DE Human cDNA encoding a novel secreted protein, Seq ID 394

KW Human; immunosuppressive; antiarthritic; ss; antirheumatic;

KW neuroprotective; antibacterial; virucide; fungicide; opthalmological;

KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;

KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;

kw skln ageing; food additive; preservative; antiproliferative.

05 Homo sapiens.

PN WO200155322-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01341.

PR 31-JAN-2000; 2000US-0179065

PR 24-FEB-2000; 2000US-0184664

PR 16-MAR-2000; 2000US-0189874

PR 18-APR-2000; 2000US-0198123

PR 07-JUN-2000; 2000US-0209467

PR 30-JUN-2000; 2000US-0215135  
PR 07-JUL-2000; 2000US-0215647

PR 07-JUL-2000; 2000US-0216880  
PR 11-THU-2000; 2000US-0217487

PR 11-JUL-2000; 2000US-0217496  
PR 14-JUL-2000; 2000US-0218290

PR 26-JUL-2000; 2000US-0220963  
PR 26-JUL-2000; 2000US-0220964

PR 14-AUG-2000; 2000US-0224518  
PR 14-AUG-2000; 2000US-0224519

PR 14-AUG-2000; 2000US-0225213

PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0231080.  
 PR 08-SEP-2000; 2000US-0231081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0233397.  
 PR 14-SEP-2000; 2000US-0233398.  
 PR 14-SEP-2000; 2000US-0233399.  
 PR 14-SEP-2000; 2000US-0233400.  
 PR 14-SEP-2000; 2000US-0233401.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 14-SEP-2000; 2000US-0233066.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235835.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239335.  
 PR 13-OCT-2000; 2000US-0239337.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251890.  
 PR 08-DEC-2000; 2000US-0251899.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 P-PSDB; AAU16228.  
 WPI: 2001-488783/53.  
 PT New nucleic acid molecules encoding 461 human secreted proteins for  
 diagnosing, preventing, treating or ameliorating medical conditions and  
 used as food additives or preservatives -  
 XX  
 PS Claim 1; SEQ ID No 394; 980bp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. Antibodies to the proteins can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
 CC include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischemia, angiodysplasia, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular disorders e.g. corneal infection, and many other  
 CC disorders listed in the specification. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to

CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence encodes a novel secreted protein of the invention.

Query Match 91.7%; Score 761; DB 22; Length 864;  
 Best Local Similarity 100.0%; Pred. No. 1e-226;  
 Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 70 GGAATACCTGGTGGAGCCCTCTTCATAGAACAGAGATGCAATCTGTGATTCA 129
DB 35 GGAATACCTGGTGGAGCCCTCTTCATAGAACAGAGATGCAATCTGTGATTCA 94
QY 130 GACCTATGAGTGCAGAGCTGAGAGCTGCTGAGAGTTTGTCAATGTCTACTACAC 189
DB 95 GACCTATGAGTGCAGAGCTGAGAGCTGCTGAGAGTTTGTCAATGTCTACTACAC 154
QY 190 CATGATTAAGCGGCGGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 249
DB 155 CATGATTAAGCGGCGGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 214
QY 250 GAATGCAATGCTGTTTCAAGACAGAAATCCTGAGTGAAGTTTGAATGTTGCTTC 309
DB 215 GAATGCAATGCTGTTTCAAGACAGAAATCCTGAGTGAAGTTTGAATGTTGCTTC 274
QY 310 CAGCGATTCACAAATCAGCGTGTGATGACGACGCTGCTGCTGCTGCTGCTGCTG 369
DB 275 CAGCGATTCACAAATCAGCGTGTGATGACGACGCTGCTGCTGCTGCTGCTGCTG 334
QY 370 CCAGACACGCTGCTGTTGTCATCTGTGATCAGTGAAGTTTGAAGGAGAACACG 429
DB 335 CCAGACACGCTGCTGTTGTCATCTGTGATCAGTGAAGTTTGAAGGAGAACACG 394
QY 430 GAGCTTCAACCAAGTCTATCTGACGCGCCAGGCTTACCCAGACACAGTGTGAA 489
DB 395 GAGCTTCAACCAAGTCTATCTGACGCGCCAGGCTTACCCAGACACAGTGTGAA 454
QY 490 GATCGCAAGTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 549
DB 455 GATCGCAAGTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 514
QY 550 TCTTGTCTTCAATCAAGCTTCTGATAGAGAAATGCAACCTGACCTCAAGAGT 609
DB 515 TCTTGTCTTCAATCAAGCTTCTGATAGAGAAATGCAACCTGACCTCAAGAGT 574
QY 610 GAGGAAACAGTATTTCTGTTGTTGGGAGACAGTGCACACTGTCGCGAGT 669
DB 575 GAGGAAACAGTATTTCTGTTGTTGGGAGACAGTGCACACTGTCGCGAGT 634
QY 670 TGAATCTTTTGTGTCAGTCTAGAGTCCCTTCCGATATATATCTTTTGT 729
DB 635 TGAATCTTTTGTGTCAGTCTAGAGTCCCTTCCGATATATATCTTTTGT 694
QY 730 CATAGTCTTTTCAAGTGAAGTATTTCTATTTTCTACTTCCAGAGAGACTC 789
DB 695 CATAGTCTTTTCAAGTGAAGTATTTCTATTTTCTACTTCCAGAGAGACTC 754
QY 790 TGATTCGGAATCTGCAAAATATATATATATATATATATATATATATATAT 830
DB 755 TGATTCGGAATCTGCAAAATATATATATATATATATATATATATATATAT
  
```

RESULT 3  
 AAS58525  
 ID AAS58525 standard; cDNA: 318 BP.  
 XX AAS58525;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX

DE cDNA #1201 encoding portion of a human colon tumour protein.  
 XX  
 KW Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200173027-A2.  
 XX  
 PD 04-OCT-2001.  
 XX  
 PF 22-MAR-2001; 2001MO-US09246.  
 XX  
 PR 24-MAR-2000; 2000US-191597P.  
 XX  
 PR 04-MAY-2000; 2000US-202024P.  
 XX  
 PR 05-MAY-2000; 2000US-202189P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Meagher MJ, Xu J, King GE;  
 XX  
 DR WPI; 2001-611627/70.  
 XX  
 PT New colon tumour proteins and related nucleic acid, useful for  
 XX treatment, prevention, diagnosis and monitoring of cancer.  
 XX  
 PS Claim 4; Page 246; 299pp; English.

Th present invention relates to the isolation of novel cDNA sequences  
 CC encoding for at least an immunogenic portion of human colon tumour  
 CC proteins. The sequences of the invention are useful in pharmaceutical  
 CC compositions and vaccines for the prevention and treatment of cancers  
 CC such as colon cancer. They are also useful for the diagnosis and  
 CC monitoring of such cancers. Antibodies to the colon tumour proteins  
 CC and antigen presenting cells that express polynucleotides encoding  
 CC colon tumour proteins can be used to inhibit the development of  
 CC cancers. T-cells that react specifically with colon tumour proteins  
 CC are useful for removing tumour cells from samples (e.g. blood) and  
 CC for cancer treatment. The polynucleotide sequences are also useful in  
 CC gene therapy. AAS7325-AAS58860 represent the cDNA sequences of the  
 CC invention that encode for portions of human colon tumour proteins.

Sequence 318 BP; 76 A; 82 C; 83 G; 76 T; 1 other;

Query Match 38.2%; Score 317; DB 23; Length 318;  
 Best Local Similarity 99.7%; Pred. No. 2e-88;  
 Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 147 CCGCAGAGCTGCTGAGAGTTTGTCAATGTCTACTACACCACTGATAGCGGCGC 206
DB 1 CCGCAGAGCTGCTGAGAGTTTGTCAATGTCTACTACACCACTGATAGCGGCGC 60
QY 207 GTTGTCTGCCGCTGTACATGAGGACAGCCACCTGCTGGAATGGCAATGCTGTT 266
DB 61 GTTGTCTGCCGCTGTACATGAGGACAGCCACCTGCTGGAATGGCAATGCTGTT 120
QY 267 CAGGACAGAAATCCTGAGTGAAGTTTGAATGTTGCTTCCAGAGTCCAAATCA 326
DB 121 CAGGACAGAAATCCTGAGTGAAGTTTGAATGTTGCTTCCAGAGTCCAAATCA 180
QY 327 GCGTGTAGACTGCCAGCTGTTCAATGATGAAGCAGCAAGCAGAGCGGCTTGG 386
DB 181 GCGTGTAGACTGCCAGCTGTTCAATGATGAAGCAGCAAGCAGAGCGGCTTGG 240
QY 387 TTGTCAATGTGATCAATGAAGTTTGAAGGGAACAAACAGGACTTCAACGAACT 446
DB 241 TTGTCAATGTGATCAATGAAGTTTGAAGGGAACAAACAGGACTTCAACGAACT 300
QY 447 TCATCTGAGCGCCAGG 464
DB 301 TCATCTGAGCGCCAGG 318
  
```

RESULT 4

AA022887  
ID AA022887 standard; cDNA to mRNA; 359 BP.  
XX  
AC AA022887;  
XX  
DT 29-AUG-1996 (first entry)  
XX  
DE Human gene signature HUMG04611.  
XX  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
XX  
OS Homo sapiens.  
XX  
PN W09514772-A1.  
XX  
PD 01-JUN-1995.  
XX  
PF 11-NOV-1994; 94MO-JP01916;  
XX  
PR 12-NOV-1993; 93JP-0355504.  
XX  
PA (MATS/) MATSUBARA K.  
PA (OKUB/) OKUBO K.  
XX  
PI Matsubara K, Okubo K;  
XX  
DR WPI; 1995-206931/27.  
XX  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
XX  
PS Claim 1; Page 1235-1236; 2245pp; Japanese.  
XX  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in AA019001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
XX  
SQ Sequence 359 BP; 85 A; 76 C; 75 G; 114 T; 9 other;  
XX  
Query Match 35.88; Score 297.4; DB 16; Length 359;  
Best Local Similarity 94.25; Pired. No. 2,7e-82;  
Matches 326; Conservative 0; Mismatches 15; Indels 5; Gaps 2;  
XX  
OY 490 GATCGAAGTGAATGTTCCGCTTCAGAGACTGGGCGGAGTGGGCGAGAGATC 549  
DB 1 GATCGAAGTGAATGTTCCGCTTCAGAGACTGGGCGGAGTGGGCGAGAGATC 60  
OY 550 TCTTTCATTCATCAGCCCTAGCTGTAGAGAAATGCAAACTGCAGCTCAAGATGT 609  
DB 61 TCTTTCATTCATCAGCCCTAGCTGTAGAGAAATGCAAACTGCAGCTCAAGATGT 120  
OY 610 GAGGAACACAGTTCATTTCTGTGTGGGAGACACTGCAGACTCCAGTCCGAGGT 669  
DB 121 GAGGAACACAGTTCATTTCTGTGTGGGAGACACTGCAGACTCCAGTCCGAGGT 180  
OY 670 TGAACCTTTTGTGTCGCAAGTTCAGAGACCTTTTCCATATATATCTTTTGT 729  
XX

DB 181 TGAACCTTTTGTGTCGCAAGTTCAGAGACCTTTTCCATATATATCTTTTGT 240  
OY 730 CATAGTTTCTTTCAGAGTAACTTTTCTATTTTCTACTGCCAGTAGAGACTC 789  
DB 241 CATAGTTTCTTTCAGAGTAACTTTTCTATTTTCTACTGCCAGTAGAGACTC 300  
OY 790 -TGATTCGTGAAA-----TTCTGACAAATATTAATATATACATG 830  
DB 301 TTGATTCGTGAAAATTTCTGTCGCAAGTTCAGAGACCTTTTCCATATATATCTTTTGT 846  
RESULT 5  
AA07014  
ID AA07014 standard; cDNA; 716 BP.  
XX  
AC AA07014;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA clone (5'-primer); SEQ ID NO:3849.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 1; SEQ ID 3849; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AA03166 to AA03168 and  
CC AA03163 to AA03172 represent human cDNA sequences; AA03246 to  
CC AA032893 represent human amino acid sequences; and AA03163 to AA03168  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.

Query Match	30.9%	Score 256.2	DB 22	Length 716
Best Local Similarity	74.3%	Pred. No. 2.9e-69		
Matches 321	Conservative 0	Mismatches 111	Indels 0	Gaps 0
114	CATCTGGGATTTTCAAGACATCATGTGCATGACAGCCCTGCAGACCTCTGAGAGTTTGCA	173		
115		174		
116		175		
117		176		
118		177		
119		178		
120		179		
121		180		
122		181		
123		182		
124		183		
125		184		
126		185		
127		186		
128		187		
129		188		
130		189		
131		190		
132		191		
133		192		
134		193		
135		194		
136		195		
137		196		
138		197		
139		198		
140		199		
141		200		
142		201		
143		202		
144		203		
145		204		
146		205		
147		206		
148		207		
149		208		
150		209		
151		210		
152		211		
153		212		
154		213		
155		214		
156		215		
157		216		
158		217		
159		218		
160		219		
161		220		
162		221		
163		222		
164		223		
165		224		
166		225		
167		226		
168		227		
169		228		
170		229		
171		230		
172		231		
173		232		
174		233		
175		234		
176		235		
177		236		
178		237		
179		238		
180		239		
181		240		
182		241		
183		242		
184		243		
185		244		
186		245		
187		246		
188		247		
189		248		
190		249		
191		250		
192		251		
193		252		
194		253		
195		254		
196		255		
197		256		
198		257		
199		258		
200		259		
201		260		
202		261		
203		262		
204				

XX MPI: 2000-256642/22.  
DR P-PSDB; AAY82323.

New human protein transport-associated polypeptide and polynucleotide useful for diagnosis, prevention and treatment of cell proliferative and secretory disorders such as leukemia, cystic fibrosis

PS Claim 9; Page 74; 75pp; English.

XX  
XX AAA08035 to AAA08042 encode the human protein transport-associated  
CC molecules (PRAMs) given in AY82317 to AY82324. The PRAMs have  
CC cytoskeletal, antiarthritic, antidiabetic, immunosuppressant,  
CC antihemorrhagic, antiallergic, antidiabetic, antipneumatic,  
CC antirheumatic, osteopathic, dermatological and antihiv activities, and  
CC hepatotropic, antigenot, antiinflammatory and antigonists are useful for  
CC regulate protein transport. PRAM proteins and antagonists are useful for  
CC preventing or treating a disorder associated with decreased or increased  
CC expression or activity of PRAM. PRAM polynucleotides are useful for  
CC diagnosing conditions associated with PRAM, comprising detecting PRAM by  
CC forming a hybridisation complex, preferably after PCR amplifying the  
CC biological sample. Diseases prevented, treated or diagnosed include cell  
CC proliferative disorders such as cancers, immune disorders, secretory  
CC disorders and other conditions associated with abnormal vesicle  
CC trafficking, such as allergies, asthma, urticaria and autoimmune  
CC haemolytic anemia. Anti-PRAM antibodies may be used as antagonists, as  
CC a targeting or delivery mechanism for bringing pharmaceutical agents  
CC into cells or tissues expressing PRAM and for diagnosis of PRAM-related  
CC disorders. PRAM, its catalytic or immunogenic fragments are useful for  
CC drug screening using libraries of compounds. PRAM polynucleotides are  
CC useful for generating hybridisation probes useful in mapping the  
CC naturally occurring genomic sequences.

XX Sequence 1146 BP; 331 A; 212 C; 228 G; 375 T; 0 other;

SQ

Query Match 30.3%; Score 251.6; DB 21; Length 1146;  
Best Local Similarity 72.4%; Pred. No. 1e-67;  
Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0

OY 114 CATCTGTGATTTTCAGACCTATGTGGATCAGGCCTCGACAGAGTGTGAGAGTTGTCA 173  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
162 CGCTGTGGATTGTTTTAAACTTAATGAATGATGAGCATGTAGAGTGCTGAGAGTTGTCA 221  
OY 174 ATGTCTACTACACACCACCATGATAGCGGGCGGCTTGTGCCCGCTGTACATGGCA 233  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
222 ATTATTACTATGAGACAATGGATGATAAAGAAGGGGACATAACACAGCTGTATCTGGACA 281  
OY 234 CAGCCACCCCTGTCTGTGATGCAATGCTGTTCAGACAGAAATCTTGAGTAGATTTT 293  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
282 AGGCCACCTTAATATGGAATGAGAAATGCTGTTCAGGGCTGGATGCCCTTAATAATTTT 341  
OY 294 TTGAATGTGTGCTTCCAGGAGATTCCAATCAGCGGTGATGAGATGGCAGCCTGTTCATG 353  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
342 TTACACATTGCTTTAGTAGAATTTCCAGGTCAATATGTTAGTTGCAACCAAGTTTCATG 401  
OY 354 ATGAGCACACACAGCAGCAGACAGCGTCTTGTGTGATCTGTGATCAGTGAAGTTTG 413  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
402 AGCAAGCACTCAGTCCCAACACTACAGATTCTTGTGTGACACAGTGAACGTGAAGTTTG 461  
OY 414 AGGGGAGCAACAACGGGACTTCAACCAACTTCACTGTGACGCCAGCCAGCCTCACCA 473  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
462 ATGGAACAACAACAATTTCTTCACACAGAACTCCGTGTGAGTGTGCTCACTCACGCCA 521  
OY 474 GCAACACAGTGTGGAAGATGTGCAAGTACAGTTCGGCTTCAGAGATGGGGCAGGTAGT 533  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
522 ACAATATGTGTGTGAATATTGCAAGTATGCTCCGTTTTCANAGATTGGTCTATGATGTT 581  
OY 534 GGGGTGGCAGAGTGTCTTTTGTGCTTCAATC 563  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
582 AAAGGGCAAAAGTCATCTCATTTGGTGC 611

RESULT 7

ABA05754  
ID ABA05754 standard; cDNA; 2581 BP.  
XX  
AC ABA05754;  
XX  
DT 04-MAR-2002 (first entry)  
XX  
DE Human NTF2 associated protein 16 coding sequence.  
XX  
KW Human; NTF2 associated protein 16; cancer; HIV infection; gene therapy;  
KM ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 147..575  
FT /tag= a  
FT /product= "NTF2 associated protein 16"  
XX  
PN CNI33332-A.  
XX  
PD 19-SEP-2001.  
XX  
PF 15-MAR-2000; 2000CN-0114918.  
XX  
PR 15-MAR-2000; 2000CN-0114918.  
XX  
PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
XX  
PI Mao Y, Xie Y;  
XX  
DR WPI: 2002-042199/06.  
DR P-PSDB; ABB04459.  
XX  
PT Polypeptide-human NTF2 associated protein 16 and polynucleotide  
XX encoding it -  
XX  
PS Claim 6; Page 24-25(Disclosure); 32pp; Chinese.  
XX  
CC The present invention provides the protein and coding sequences of human  
CC NTF2 associated protein 16. The sequences can be used in the treatment of  
CC cancer and HIV infection, as well as other diseases. The present sequence  
CC is the coding sequence of the invention.  
XX  
SQ Sequence 2581 BP; 791 A; 449 C; 452 G; 889 T; 0 other;  
XX  
Query Match 30.3%; Score 251.6; DB 24; Length 2581;  
Best Local Similarity 72.4%; Pred. No. 1.7e-67;  
Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
XX  
QY 114 CATCTGATTTCAAGACCTATGTGATCAGGCTGACAGCTGCTGAGAGTTTGTCA 173  
DB 154 CGCTCTGATTTTAAACTTATGATGATCAGCAGATGATGCTGAGAGTTTGTCA 213  
QY 174 ATGTCTACTACACACCATGATGATGAGGCGGCTTGTGCTCCGCTGTACATGGCA 233  
DB 214 ATTTTACTATGAGCAATGATGATGAGGCGGCTTGTGCTCCGCTGTACATGGCA 273  
QY 234 CAGCCACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 293  
DB 274 AGGCCACCTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 333  
QY 294 TTGAATGTTGCTTCCAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 353  
DB 334 TTGACATATGCTTCTAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 393  
QY 354 ATGAACCCAC 413  
DB 394 AGCAACCACTAGTCTCCCAACACACACACACACACACACACACACACACACACAC 453  
QY 414 AGGGGAAACAAACAGGAGCTTCAACACACACACACACACACACACACACACACAC 473  
DB 454 ATGGAACAAACACACATTTCTTCAACACACACACACACACACACACACACACAC 513

QY 474 GCAACACAGTGTGAGAGATGCAAGTACTGCTTCCAGTCTGCAAGACTGGCCAGCTACT 533  
DB 514 ACAATAGTGTGTGAGATGCAAGTACTGCTTCCAGTCTGCAAGACTGGCTAGTACT 573  
QY 534 GGGGTGTCAGAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 563  
DB 574 AAGGGGCAAAAGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 603  
XX  
RESULT 8  
ID AAH17938  
XX AAH17938 standard; cDNA; 2692 BP.  
XX  
AC AAH17938;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human cDNA sequence SEQ ID NO:17706.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.  
XX  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
DR WPI: 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
PS Claim 8; SEQ ID 17706; 2537pp + CD ROM; English.  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any special methods. AAH03166 to AAH13658 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX

OY	114	CATGTGGGAAATTTAAAGACCTATGGAATCAGGCGCTGCAGAGACTCTGAGAGATTGTCA	173
Db	275	CGTCTTGGAATTTAAACTTATGTAGTATCAGGCATGTAGAGCTGCTGAGAGATTGTCA	334
OY	174	ATGTCTACTACACCAACCATGGATTAAGCGGGCGCTTTGCTGTCCCGCTGTACATGGCA	233
Db	335	ATATTTCATAGACACATGGATTAAGAGACGGGCACTAACCAAGCTGTATCTGACA	394
OY	234	CAGCCACCTGCTGTGGAATGGCAATGCTGTTTCAGGACAGAAATCCTTGAATGATTT	293
Db	395	AGGCCACCTTATATGTGAATGGAATGCTGTTTCAGGGCTGGATGCCCTAAATATTTT	454
OY	294	TTGAAATGTGCTTCCACGCGAGTTCCAAAATCAGGCTGGTAGACTGCCAGCCTGTTCATG	353
Db	455	TTGACACACTTCCTTCTGTAGTAGTTCAGGTCATATATGTAGATTGGCAACAGTTTCAG	514
OY	354	ATGAAGCCACACACAGCCAGACAGACAGGTCCTTGTGTCAATCTGTGGATCAGTAAATTG	413
Db	515	AGCAAGCAACACAGTCCCAACTACAGTTCTGTGTGACAGGTGGAACCTGTGAAGTTG	574
OY	414	AGGGGACAAACCAACGGGACTTCAACACAGAACTTCAATCTGACCGCCAGGCTCACCCA	473
Db	575	ATGGAACAAACAAACATTTCTTCAACACAGAACTTCTGCTGACTGTCAGTCCACTCCCA	634
OY	474	GCACACAGTGGGAAGATCCGAAGTGAAGTCTGCTCCGCTCCAGAGCTGGGCCAGCTAGT	533
Db	635	ACAAATACGTGTGGAGATGCAAGTGAATGCTTCGTTTCCAGATTTGGTCTAGTAGTT	694
OY	534	GGGGGTGGCAGAGTCTCTTTGCTTCATTC	563
Db	695	AAAGGGCAAAAGTCCATTCATTTGGT	724

RESULT 9  
ABV24535  
ID ABV24535 standard; cDNA; 2994 BP.

XX	AC	ABV24535;
XX	DT	16-SEP-2002 (first entry)
XX	XX	
XX	XX	Human prostate expression marker; cDNA 24526.
XX	KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX	OS	pharmacogenomic marker; gene; ss.
XX	PN	Homo sapiens.
XX	PD	WO200160860-A2.
XX	PF	23-AUG-2001.
XX	PR	20-FEB-2001; 2001WO-US05171.
XX	PR	17-FEB-2000; 2000US-183319P.
XX	PR	16-MAR-2000; 2000US-189862P.
XX	PR	25-MAY-2000; 2000US-207454P.
XX	PR	09-JUN-2000; 2000US-211314P.
XX	PR	18-JUL-2000; 2000US-219007P.
XX	PR	13-DEC-2000; 2000US-255281P.
XX	PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX	PI	Schlegel R, Endege WO, Monahan JE;
XX	DR	WPI; 2001-662795/76.
XX		

Oy		114	CATCTGTGGATTTCGAAGCCCTTGTGTAATCGACGCCCTGCAGAGCCTCCTGAGGGATTTGCA	173
Db		116	CCTCTCTGGATTTTAAACTTATGTAGATCAGAGCATGTAGACCTGCTGAGGAGTTTGCA	175
Oy		174	ATGCTACTACCAACCACCATGATTAAGCGCGCCGTTTGGTGTCCCGCCCTGTACATGGGCA	233
Db		176	ATATTACTATGAGCAATAGATTAATAAAGAAGACGGGCATTAACCAAGGCTGTATCTGGACA	235
Oy		234	CAGCCACCCCTGGTCTGGAATGGCAATGCTGTTTCAGAGCAAGAACAAATCCTTAGCTAGTTT	293
Db		236	AAGCCACCTTAATATGGAAATGGAAATAGCTGTTTCAGGGCTGGATGGCCCTAAATATTTT	295
Oy		294	TTGAAATGTTGCCCTTCCACGAGTTCCAATATAGCGTGGTAACTGCCAGCCTGTTCATG	353
Db		296	TTCACACATTTGCCCTTCTAGTGAAGTTCCAGGTCAATATGTAAATTTCCACACAGTTTCATG	355
Oy		354	ATGAAGCCACACCAACCAAGCACAGACCGGCTCTGTTGTCACTGTGTGGATCAGTGAATTGG	413
Db		356	AGCAACACATCTAGCTGCCCAACTACAGTTCTTGTGTGACAGAGTGAACGTGMAATTGG	415
Oy		414	AGGGGAACAACAACGGGACTTCACCAAGACTTCATCTGACCGCCAGGCTCAACCCA	473
Db		416	ATGGAACAACAACAATTTCTTCAACACAGAACTCTCGTGAAGTCTCAGTCCACATCCCA	475
Oy		474	GCAACACAGTGTGGAAGATTCGCAAGTACCTCGCTCCGGTCCAGAGCTGGGCCAGGCTAGT	533
Db		476	ACAAATCTGTGTGGAAGATTCGAAGATGCTTCGCTTTCAAGATTTGGTCTAGTAGTT	535
Oy		534	GGGGGTGGCAGAGCTCTTTGCTTTGTTTATTC	563
Db		536	AAAAGGGCAAAAGTCCATTCTCATTTGGTTC	565
RESULT 10				
AA\$26214	ID	AA\$26214 standard; cDNA; 1387 bp.		
XX	AA\$26214;			
XX	AC			
XX	DT	07-NOV-2001 (first entry)		
XX	DE	Human cDNA encoding a novel secreted protein, Seg ID 393.		
XX	KM	Human: immunosuppressive; antiarthritic; ss; antirheumatic;		
XX	KW	-cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;		







PR 01-NOV-2000; 200005-245084P.  
XX  
XX (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI: 2002-188264/24.  
XX  
PT Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set  
XX  
PS Claim 1; SEQ ID 890; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilms' tumor.  
XX  
SQ Sequence 247 BP; 89 A; 41 C; 45 G; 72 T; 0 other:  
Query Match 26.9%; Score 223; DB 24; Length 247;  
Best Local Similarity 100.0%; Pred. No. 3.6e-59;  
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 608 GTGAGACACACAAAGTTCATTTCTGTTGTCGGAGACACTGCAGACTCCACTGTCGGAG 667  
DB 247 GTGAGACACACAAAGTTCATTTCTGTTGTCGGAGACACTGCAGACTCCACTGTCGGAG 188  
QY 668 GTTGAAGCTCTTTTGTGCTCAAGTTCAGAGTCCCTTCCGAATATATATCTGTT 727  
DB 187 GTTGAAGCTCTTTTGTGCTCAAGTTCAGAGTCCCTTCCGAATATATATCTGTT 128  
QY 728 GTCATAGTTTCCCTTTCAAGTAGTAACTTTCTATTTTCTACTTGCCCACTAGAGAC 787  
DB 127 GTCATAGTTTCCCTTTCAAGTAGTAACTTTCTATTTTCTACTTGCCCACTAGAGAC 68  
QY 788 TCTGATTCGGAATTCGACAAATATATTAATATATACATG 830  
DB 67 TCTGATTCGGAATTCGACAAATATATTAATATATACATG 25  
RESULT 12  
AA10492/c  
ID AA10492 standard; DNA; 133 BP.  
XX  
XX AA10492;  
XX  
DT 30-MAR-1999 (first entry)  
XX  
DE Human biallelic polymorphic DNA fragment WI-13859.  
XX  
XX Polymorphism: biallelic; human; forensic; paternity testing; disease;  
XX detection; phenotypic typing; characteristic; infection; hereditary;  
XX autoimmune disease; cancer; inflammation; drug; therapy; medication;  
XX treatment; marker; ss.  
XX  
OS Homo sapiens.  
XX

PN MO9820165-A2.  
XX  
XX 14-MAY-1998.  
PD  
XX 05-NOV-1997; 97MO-0520313.  
PE  
XX 06-NOV-1996; 96US-0030455.  
PR  
XX  
XX (WHD) WHITEHEAD INST BIOMEDICAL RES.  
PA  
XX Hudson T, Lander ES, Wang D;  
PI  
XX  
XX WPI: 1998-286974/25.  
XX  
PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
XX  
PS Claim 1; Page 55; 310pp; English.  
XX  
CC AAX10269-X12937 are human DNA fragments which contain biallelic  
CC polymorphic markers which have been isolated using the primers  
CC represented in AAX09121-X10268. The base occupying the polymorphic site  
CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments  
CC can be used in methods for determining polymorphic forms in an individual  
CC for use in e.g. forensics, paternity testing or for phenotypic typing for  
CC diseases such as agammaglobulinemia, diabetes insipidus, Lasch-Nyhan  
CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,  
CC familial hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC hemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
XX  
SQ Sequence 133 BP; 54 A; 20 C; 24 G; 34 T; 1 other:  
Query Match 16.0%; Score 132.6; DB 19; Length 133;  
Best Local Similarity 99.2%; Pred. No. 4.1e-31;  
Matches 132; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 682 TGTGCTCAAGTCTAGAGTCCCTTCTCGAATATATATCTGTTGTCATAGTTTCCT 741  
DB 133 TGTGCTCAAGTCTAGAGTCCCTTCTCGAATATATATCTGTTGTCATAGTTTCCT 74  
QY 742 TTCAAAGTAACTTTCTATTTTCTACTTGCCCACTAGAGACTGATTCGAAA 801  
DB 73 TTCAAAGTAACTTTCTATTTTCTACTTGCCCACTAGAGACTGATTCGAAA 14  
QY 802 TTCTGACAAATAA 814  
DB 13 TTCTGACAAATAA 1  
RESULT 13  
AAS26708/c  
ID AAS26708 standard; DNA; 8918 BP.  
XX  
XX AAS26708;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human genomic DNA encoding partial novel secreted protein, Seq ID 1682.  
XX  
XX Human; immunosuppressive; antiarthritic; ds; antirheumatic;  
XX cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;  
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
XX vulnerable; secreted protein; rheumatoid arthritis;

KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin aging; food additive; preservative; antiproliferative.  
XX Homo sapiens.  
OS  
FN MO200155322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-USO1341.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226866.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231245.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0233397.  
PR 14-SEP-2000; 2000US-0233398.  
PR 14-SEP-2000; 2000US-0233399.  
PR 14-SEP-2000; 2000US-0233400.  
PR 14-SEP-2000; 2000US-0233401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244611.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246603.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251866.  
PR 08-DEC-2000; 2000US-0251869.

PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-488783/53.  
XX  
XX New nucleic acid molecules encoding 461 human secreted proteins for  
XX PT diagnosing, preventing, treating or ameliorating medical conditions and  
XX PT used as food additives or preservatives -  
XX  
XX PS Disclosure; SEQ ID NO 1682; 980pp; English.  
XX  
XX The invention relates to isolated nucleic acid molecules and their  
XX CC encoded secreted proteins. The nucleic acids and proteins are used to  
XX CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
XX CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
XX CC are also used in diagnosing a pathological condition or susceptibility  
XX CC to a pathological condition. Antibodies to the proteins can also  
XX CC be used in alleviating symptoms associated with the disorders and in  
XX CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
XX CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
XX CC include autoimmune diseases e.g. rheumatoid arthritis, breast or liver,  
XX CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
XX CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
XX CC e.g. cerebral ischaemia, angiodysplasia, nervous system disorders e.g.  
XX CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
XX CC and ocular disorders e.g. corneal infection, and many other  
XX CC disorders listed in the specification. The polypeptides can also  
XX CC be used to aid wound healing and epithelial cell proliferation, to  
XX CC prevent skin aging due to sunburn, to maintain organ before  
XX CC transplantation, for supporting cell culture of primary tissues, to  
XX CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
XX CC as a food additive or preservative to increase or decrease storage  
XX CC capability, fat content, lipid, protein, carbohydrate, vitamins,  
XX CC minerals, cofactors and other nutritional components. The present  
XX CC sequence is a genomic DNA encoding a partial novel secreted protein of  
XX CC the invention.

Query Match 13.6%; Score 113; DB 22; Length 8918;  
Best Local Similarity 70.0%; Pred. No. 6.1e-24;  
Matches 152; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 347 GTTCATGATGACCAACACCAAGCAGACAGCAGCTCTTGTTCATCTGGATGATG 406  
DB 2199 GATGAGAGCAACCACTGCTCCCAACATGATCTTGTGACCACTGGACTGTG 2140  
QY 407 AACTTTAGAGGGAACAACAACGGAATTCACACAGACTTCATCTGACCGCCAGCC 466  
DB 2139 AACTTTAGAGGGAACAACAACATTTCTTAACCAAGACTTCTGCTGACTGCTAC 2080  
QY 467 TCACCAAGCAACAGTGTGAGATGCAAGTGTCTCCGCTTCCAGACTGGCC 526  
DB 2079 ACTCCCAACATCTGCTGTGAGATGCAAGTGTCTCCGCTTCCAGATGTGCT 2020  
QY 527 AGCTAGTGGGGGTGAGAGAGTCTCTTGTCTCATTC 563  
DB 2019 AGTAGTTAAAGGGGCAAAAGTCCATCTCATTTGGTTC 1983

RESULT 14  
ID AAS26709/c  
XX AAS26709 standard; DNA; 8919 BP.  
XX  
XX AAS26709;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
XX Human genomic DNA encoding partial novel secreted protein, seq ID 1683.

XX  
XX Human; immunosuppressive; antiarthritic; ds; antirheumatic;  
XX KW cytostatic; cardiac; vasotropic; cerebroprotective; nootropic;  
XX KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
XX KW valvular; secreted protein; rheumatoid arthritis;  
XX KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
XX KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
XX KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
XX KW corneal infection; wound healing; epithelial cell proliferation;  
XX KW skin ageing; food additive; preservative; antiproliferative.  
XX  
XX Homo sapiens.  
XX  
XX W0200155322-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX  
XX 17-JAN-2001; 2001WO-US01341.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 14-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226868.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.

RESULT 15  
AAS75617  
ID AAS7  
XX

AC AAS75617;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #11421.

XX Human: chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX MO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG11430.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping; identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX Claim 1; SEQ ID No 11421; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (I) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1035 BP; 318 A; 219 C; 222 G; 276 T; 0 other;

XX Query Match

XX Best Local Similarity 12.7%; Score 105.2; DB 23; Length 1035;  
 XX Matches 134; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

XX 114 CACTGTGATTCACAGACTATGTGATGAGCCCTGCACAGACTGCTGAGAGTTTGTCA 173

XX 202 CGTCTGTGATTTTAAACTATGATGATGAGCATGATGAGCTGCTGAGAGTTTGTCA 261

XX 174 ATGTCTACTACACACCATGATGATGAGCGGCGGCTTGTCTCCGCTGTACATGGCA 233

XX 262 ATATTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 321

XX 234 CAGCCACCCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 293

XX 322 AGGCCACCTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 381

QY 294 TT 295  
 DB 382 TT 383

Search completed: June 28, 2003, 01:09:32  
 Job time : 215.801 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:24:04 ; Search time 1923.73 Seconds

(Without alignments)  
9647.960 Million cell updates/sec

Title: US-09-763-902b-15

Perfect score: 1146

Sequence: 1 taccctcgagagagagatggg.....tttaagtaattttgtttt 1146

Scoring table: IDENTITY\_NUC  
Gapex 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_estchun:\*  
3: em\_estcin:\*  
4: em\_estlmu:\*  
5: em\_estmuv:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	847	73.9	934	9	AL539324
2	834	72.8	978	9	AL548271
3	824.4	71.9	993	14	BO051357
4	789.4	68.9	793	9	AL598693
5	775	67.6	881	11	AF212223
6	711.2	62.1	781	9	AU123481

7	706.2	61.6	770	9	AU137118
8	682.8	59.6	745	12	BE887056
9	681.4	58.5	782	13	BI458206
10	666.8	58.2	721	10	AV114243
11	656.4	57.3	743	13	BI761278
12	649.2	56.6	819	13	BI597498
13	612.2	53.4	878	12	BC035698
14	611	53.3	629	12	BF080705
15	611	53.3	629	13	BE995586
16	610.4	53.3	713	10	AV118042
17	587.4	51.3	884	12	BF205046
18	579.4	50.6	750	12	BE718308
19	562.4	49.1	666	10	BE390803
20	560.8	48.9	805	12	BE204966
21	558	48.1	702	13	BI828542
22	551.6	48.1	996	12	BE480012
23	541	47.2	931	12	BE717309
24	526	45.9	526	14	BM833679
25	523	45.6	1016	10	BE540322
26	511	44.6	511	9	AA400543
27	503.4	43.9	506	13	BM147972
28	502.2	43.8	574	12	BE004693
29	491.2	42.9	716	9	AU133525
30	473.6	41.3	493	9	AL702205
31	471	41.1	568	10	AV716990
32	448.6	39.1	487	12	BF087098
33	436	38.0	444	9	AA831754
34	402.6	35.1	467	10	AV763565
35	395.6	34.5	575	10	AV757524
36	395.6	34.5	625	10	BB622341
37	393.2	34.3	800	13	BI079409
38	390.6	34.1	736	14	BM963180
39	365.4	31.9	656	10	BE287040
40	363	31.7	458	14	N27567
41	355.2	31.0	521	12	BF087064
42	355.2	31.0	521	13	BE995575
43	352.8	30.8	421	10	AV713168
44	346.6	30.2	560	9	AI526545
45	342.6	29.9	561	9	AI956595

## ALIGNMENTS

RESULT 1  
AL539324  
LOCUS  
DEFINITION AL539324 L1 FL013 Fbri1 Homo sapiens cDNA clone CS0DF034YD22 5  
PRIME, mRNA sequence.  
ACCESSION AL539324  
VERSION AL539324.1 GI:12868435  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT  
Contact: Genoscope  
Genoscope Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers  
1. 934  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DF034YD22"  
/dev\_stage="L1 FL013 Fbri1"  
/dev\_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"  
/lab\_host="DH10B"



/note="Organ: Fetal brain; Vector: pcwvSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcwvSPORT 6 vector. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: [filang@lifetech.com](mailto:filang@lifetech.com) URL: <http://fulllength.invitrogen.com>"

BASE COUNT 269 a 170 c 180 g 304 t 11 others  
ORIGIN  
Query Match 73.9%; Score 847; DB 9; Length 934;  
Best Local Similarity 95.9%; Pred. No. 2e-202;  
Matches 892; Conservative 4; Mismatches 31; Indels 3; Gaps 3;  
QY 138 CTGCTACAAAGTCCAGATGCGCCAGCTCTCTGATTTTAAACCTATGATGATCAGGCAT 197  
DB 3 CTGGGAAATAGCGAGAGCGGGGATGTATGGGATTTTAAACCTATGATGATCAGGCAT 62  
QY 198 GTAGAGCTGCTGAGAGTGTCTCATATTTACTATGAGCATGATGATAAAGACGGG 257  
DB 63 GTAGAGCTGCTGAGAGTGTCTCATATTTACTATGAGCATGATGATAAAGACGGG 122  
QY 258 CACTAACAGGCTGATCTGACAAAGGCCACCTTAATGAAATGGAATGCTGTTTCAG 317  
DB 123 CACTAACAGGCTGATCTGACAAAGGCCACCTTAATGAAATGGAATGCTGTTTCAG 182  
QY 318 GGCCTGATGCCCTAATAATATTTTGTGACACATGCTCTGATGATGATGATGATGATGAT 377  
DB 183 GGCCTGATGCCCTAATAATATTTTGTGACACATGCTCTGATGATGATGATGATGATGAT 242  
QY 378 TGTGATGATGCAACAGTCAAGTCAAGCAACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 437  
DB 243 TGTGATGATGCAACAGTCAAGTCAAGCAACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 302  
QY 438 TGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 497  
DB 303 TGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 362  
QY 498 TGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 557  
DB 363 TGTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 422  
QY 558 GTTTTCAAGTGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 617  
DB 423 GTTTTCAAGTGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 482  
QY 618 TTCCACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 677  
DB 483 TTCCACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 542  
QY 678 AAATTAATTTCTTAATATTTTCTTAATTTCTTAATTTCTTAATTTCTTAATTTCTTAATTT 737  
DB 543 AAATTAATTTCTTAATATTTTCTTAATTTCTTAATTTCTTAATTTCTTAATTTCTTAATTT 602  
QY 738 TGAGAGATTTGCTCTTAAGAGCTTTAAACATTTTGTACATGCTTTATATACATGCA 797  
DB 603 TGAGAGATTTGCTCTTAAGAGCTTTAAACATTTTGTACATGCTTTATATACATGCA 662  
QY 798 CTAAATGCTCTTAATATTTTAAACATGCTTTGATGATGATGATGATGATGATGATGATGATG 857  
DB 663 CTAAATGCTCTTAATATTTTAAACATGCTTTGATGATGATGATGATGATGATGATGATGATG 722  
QY 858 AACCCAGCTATTTGCAAAATATAATCTTTTATATATATATATATATATATATATATATATAT 917  
DB 723 AACCCAGCTATTTGCAAAATATAATCTTTTATATATATATATATATATATATATATATATAT 782  
QY 918 ATTTAATGCTCTGAG 977  
DB 783 ATTTAATGCTCTGAG 842

QY 978 ACCTGCTGCTTTTCACTGATTAACACTGTAATGCCATTTTAAATATATGCTTTTCAATC 1037  
DB 843 ACCTGCTGCTTTTCACTGATTAACACTGTAATGCCATTTT-AAATATGCTTTTTC-AAATC 900  
QY 1038 AGCTGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1067  
DB 901 AGCTGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 929  
RESULT 2  
LOCUS AL548271 978 bp mRNA linear EST 16-FEB-2001  
DEFINITION AL548271 LTI\_NFL006.PL2 Homo sapiens cDNA clone CS001016TK06 5  
prime, mRNA sequence.  
ACCESSION AL548271  
VERSION AL548271.1 GI:12883118  
KEYWORDS EST  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 978)  
AUTHORS Li W.B., Gruber C., Jesse J. and Polayes D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Genoscope - Centre National de Sequencage  
Genoscope BP 191 91006 Evry cedex - France  
Email: [seque@genoscope.cns.fr](mailto:seque@genoscope.cns.fr) Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
FEATURES  
source  
1. 978  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS001016TK06"  
/clone\_11b="LTI\_NFL006.PL2"  
/tissue="type="Placenta"  
/note="Vector: pcwvSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pcwvSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: [filang@lifetech.com](mailto:filang@lifetech.com) URL: <http://fulllength.invitrogen.com>"

BASE COUNT 288 a 191 c 194 g 303 t 2 others  
ORIGIN  
Query Match 72.8%; Score 834; DB 9; Length 978;  
Best Local Similarity 96.8%; Pred. No. 3.8e-199;  
Matches 902; Conservative 2; Mismatches 22; Indels 6; Gaps 5;  
QY 41 TTTGGGGGTTTGGCTCTCTTAT 99  
DB 9 TTGGAGGGTTTGGCTCTCTTAT 68  
QY 100 ACCTGAGAGAGTATGAGAGC -CGACACTGCCAGACACTGCTACAAAGTCCAGATG 158  
DB 69 ACCTGAGAGAGTATGAGAGCAGACACTGCCAGACACTGCTACAAAGTCCAGATG 128  
QY 159 CGACGCTGATTTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 216  
DB 129 CGACGCTGATTTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 188  
QY 217 TGTCAATATTTACTAT -GAGACATGATGATGATGATGATGATGATGATGATGATGATGATG 275  
DB 189 AGTCAATATTTACTATGAGACAAATGATGATGATGATGATGATGATGATGATGATGATGATG 248  
QY 276 TGACAGAGCCACTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 335  
DB 249 TGACAGAGCCACTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 308

OY		336	ATTATTTTGCACATTCGCTTCATGTAGTGATCCAGGCTCAATAATGTAAATTGCCAACCG	395
Db		309	TTTTTTTTGCACATTCGCTTCATGTAGTGATCCAGGCTCAATAATGTAAATTGCCAACCG	368
OY		396	TTCATGAGCAGCAACTCAGTCCCAAACACTACAGTCTTGTTGTGACACAGTAGTAAGTGA	455
Db		369	TTCATGAGCAAGCAACTCAGTCCCAAACACTACAGTCTTGTTGTGACACAGTAGTAAGTGA	428
OY		456	AGTTTGTGGAAACAACAACATTTCTTCAACGAGAATTCCTGCTGACTGCTAGTCCA	515
Db		429	AGTTTGTGGAAACAACAACATTTCTTCAACGAGAATTCCTGCTGACTGCTAGTCCA	488
OY		516	CTCCCACACATACCTGTGTGGAAGATTTGCAAGTGTATGCTTCCGTTTAAATTTGTGTA	575
Db		489	CTCCCACACATACCTGTGTGGAAGATTTGCAAGTGTATGCTTCCGTTTCAATTTGTGTA	548
OY		576	GTAAGTAAAG-GGAGCAAAAGCCATTCGATTTGGTCCATAGTTCCAGCAATTGAATT	634
Db		549	GTAAGTAAAGAGGSCAAAAGCCATTTCTCATTTGGTCCATAGTTCCAGCAATTGAATT	608
OY		635	TATGTGAATTAATTTTGAATGTAGAAGCACATTAATATGTGCTGAACATAATTTCTTTAA	694
Db		609	TATGTGAATTAATTTTGAATGTAGAAGCACATTAATATGTGCTGAACATAATTTCTTTAA	668
OY		695	TATTTTCATTCGCTGACGACACCTTTCTTACAGCTGCGACGTTGGAGCAATTCGCCCTCA	754
Db		669	TATTTTCATTCGCTGACGACACCTTTCTTACAGCTGCGACGTTGGAGCAATTCGCCCTCA	728
OY		755	AGACCTTTAAACATATTTTATTTTACATGCGCTTATATACATTCACATTAATGACATTTCTTATA	814
Db		729	AGACCTTTAAACATATTTTATTTTACATGCGCTTATATACATTCACATTAATGACATTTCTTATA	788
OY		815	ATAATATTAAACACATGATCTTGTGTACTAACATACTCACTGTGAACCAGCCTATTGCCAA	874
Db		789	ATAATATTAAACACATGATCTTGTGTACTAACATACTCACTGTGAACCAGCCTATTGCCAA	848
OY		875	AAAAAATATCTTTTATTAATATTAATATCTATGGAGATGTCAGCAATAATTAACACTGTGGGAA	934
Db		849	AAAAAATATCTTTTATTAATATTAATATCTATGGAGATGTCAGCAATAATTAACACTGTGGGAA	908
OY		935	GAACTGGAGATTTTGTGTTATAGGTATATT 966	
Db		909	AGAAGTGGAGATTTTGTGTTATAGGTATATT 940	
RESULT 3				
B0051357		993 bp	mRNA	linear EST 29-MAR-2007
LOCUS				
DEFINITION:	AGNCNCURF_6763695 NIH_MGC_68 Homo sapiens CDNA clone IMAGE:5792758			
ACCESSION	B0051357			
KEYWORDS	B0051357.1 GI:19810697			
SOURCE	EST.			
ORGANISM	human.			
	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Cnielata; Vertebrate; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 993)			
AUTHORS	NIH-MGC http://mgi.nci.nih.gov/.			
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@bbs.femail.nih.gov Tissue Procurement: DCD/D/P/gazdar cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LAM12891 row: j column: 23 High quality sequence stop: 717. Location/Qualifiers			
FEATURES				

```

source
1. 993
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5792758"
/clone_1ib="NIH-MGC_68"
/tissue="type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: PCW-SPORE6; Site: 1.4 NotI; Site: 2: SalI; Cloned unidirectionally. Primer: O190 dt. Average insert size 1.8 kb. Library constructed by Life Technologies."
BASE COUNT      306 a      174 c      192 g      321 t
ORIGIN
Query Match      71.9%; Score 824.4; DB 14; Length 993;
Best Local Similarity 97.2%; Pred. No. 18-196;
Matches 861; Conservative 0; Mismatches 21; Indels 4; Gaps 2

169 GGATTTTAAACTTATGTAGATCAGGCATGTAGAGCTGCTGAGAGACTTGTCTAATTTTA 228
|||||
42 GGATTTTAAACTTATGTAGATCAGGCATGTAGAGCTGCTGAGAGACTTGTCTAATTTTA 101
|||||
229 CTATGAGCAATATGATTAAGAAGAGGCGGCACTTAACAGAGCTGTATCTGGACAGGCCAC 288
|||||
102 CTATGAGCAATATGATTAAGAAGAGGCGGCACTTAACAGAGCTGTATCTGGACAGGCCAC 161
|||||
289 CTTAATATGGAATGGAATGCTGTTTCAGGCGTGGATGCCCTTAATATTTTTTTCACAC 348
|||||
162 CTTAATATGGAATGGAATGCTGTTTCAGGCGTGGATGCCCTTAATATTTTTTTCACAC 221
|||||
349 ATTGCTTCTATGAGTTCAGAGTCAATATGTAGATTGCCAACAGCTTCATGAGAGC 408
|||||
222 ATTGCTTCTATGAGTTCAGAGTCAATATGTAGATTGCCAACAGCTTCATGAGAGC 261
|||||
409 AACTGAGTCCCAACTACAGTCTGTTGTGACACATGGAACGTGAAGTTTGATGGAA 468
|||||
282 AACTGAGTCCCAACTACAGTCTGTTGTGACACATGGAACGTGAAGTTTGATGGAA 341
|||||
459 CAACAACATTTCTTCAACACAGAACTTCTGCTGACTGCTCAGTCCATCCCAACATAC 528
|||||
342 CAACAACATTTCTTCAACACAGAACTTCTGCTGACTGCTCAGTCCATCCCAACATAC 401
|||||
529 TGTGTGGAAGATTGCAAGTGATGCTTCCGCTTTTCAAGTTGGTCTAGTAAGTAAAGGG 588
|||||
402 TGTGTGGAAGATTGCAAGTGATGCTTCCGCTTTTCAAGTTGGTCTAGTAAGTAAAGGG 461
|||||
589 CAAAAGTCCATTCATTTGGTGCATTTAGTTCACAGAAATTGAATTTATGTGAATTTT 648
|||||
462 CAAAAGTCCATTCATTTGGTGCATTTAGTTCACAGAAATTGAATTTATGTGAATTTT 521
|||||
649 TGATTTGTAAGACACTATATATATGTGCTGAACATAATTTCTTTAATATTTTCTATTCCT 708
|||||
522 TGATTTGTAAGACACTATATATATGTGCTGAACATAATTTCTTTAATATTTTCTATTCCT 581
|||||
709 GTACAGACCTTTTCTAGCAGCTGCCAGTTTGGAGCATTTGCCCTCTAAGAGCTTTAAACT 768
|||||
582 GTACAGACCTTTTCTAGCAGCTGCCAGTTTGGAGCATTTGCCCTCTAAGAGCTTTAAACT 641
|||||
769 ATTTTCTTACATGCTTATATACATTCCTCACTAATGATGATCTTATATATATTTTAAAC 828
|||||
642 ATTTTCTTACATGCTTATATACATTCCTCACTAATGATGATCTTATATATATTTTAAAC 701
|||||
829 ATGATCTTGTACTAACAATCTACTGTGACCCAGCCATTTGCAAAAATATTAATCTTTT 888
|||||
702 ATGATCTTGTACTAACAATCTACTGTGACCCAGCCATTTGCAAAAATATTAATCTTTT 761
|||||
888 TATTAATATATCTATAGGATGTCAGACAAATATACACTCTGGGAAAGATGAGTTT 948
|||||
762 TATAATATATCTATAGGATGTCAGACAAATATACACTCTGGGAAAGATGAGTTT 821
|||||
949 T-GGTTTATAGGTTAATTTCTAGTAAGACAGTTGCCGTTTTCGTTAACCTGGTAA 100
|||||
822 TGGGTAATTAGGTTAATTTCTAGTAAGACAGTTGCCGTTTTCGTTAACCTGGTAA 881

```

QY 1008 TGCATTATTAAT--ATGCTTTTCAATCAGTGTGAAA 1050  
 DB 882 ATGCCATTTTAATATATATGCTTTTCAATCAGTGTGAAA 927  
 RESULT 4  
 AL598693 793 bp mRNA linear EST 14-AUG-2001  
 LOCUS DKF2P313A1822.1 313 (synonym: h1cc2) Homo sapiens cDNA clone  
 DEFINITION DKF2P313A1822.5 mRNA sequence.  
 ACCESSION AL598693.1 GI:15161384  
 VERSION EST.  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 793)  
 Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann  
 'S. EST (Duesterhoeft, et al.)  
 TITLE EST (Duesterhoeft, et al.)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Duesterhoeft A  
 MIPS  
 Am Klopferseplatz 18a D-82152 Martinsried, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing  
 consortium of the German Genome Project.  
 No si sequence available.  
 This clone (DKF2P313A1822) is available at the RZPD in Berlin:  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 FEATURES  
 source  
 1. 793  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKF2P313A1822"  
 /clone\_1lb="313 (synonym: h1cc2)"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Vector: pTriplEx2; Site\_1: sfliA; Site\_2: sfliB;  
 cDNA: collection"  
 BASE COUNT 219 a 160 c 172 g 241 t 1 others  
 ORIGIN  
 Query Match 68.9%; Score 789.4; DB 9; Length 793;  
 Best Local Similarity 99.7%; Pred. No. 6.2e-188;  
 Matches 790; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 12 GAGATGGAGAGGTTGAAAATTTTGGCTTTGGCGGTTTGGCTTCTTCAATAGTAT 71  
 DB 2 GAGAAAGGAGGAGTGAATAATTTTGGCTTTGGCGGTTTGGCTTCTTCAATAGTAT 61  
 QY 72 GATCATTTCCGACGCCCTGGGACCGGACGAGAGTAGTAGAGCCGACACTGCCAG 131  
 DB 62 GATCATTTCCGACGCCCTGGGACCGGACGAGAGTAGTAGAGCCGACACTGCCAG 121  
 QY 132 AACACACTGCTACAGAGTCCAGATGCGACGCTCTCTGATTTTAAACTTTGTAGATC 191  
 DB 122 AACACACTGCTACAGAGTCCAGATGCGACGCTCTCTGATTTTAAACTTTGTAGATC 181  
 QY 192 AGGCATGTAGAGCTGCTGAGAGAGTTTGTCAATATTACTATGAGACATGATGATAAGAA 251  
 DB 182 AGGCATGTAGAGCTGCTGAGAGAGTTTGTCAATATTACTATGAGACATGATGATAAGAA 241  
 QY 252 GAGCGGCACTATACGAGGCTGTATCCGAGAGGACCCCTAATATGAGATGAAATGCTG 311  
 DB 242 GAGCGGCACTATACGAGGCTGTATCCGAGAGGACCCCTAATATGAGATGAAATGCTG 301  
 QY 312 TTTCAGGCGCTGATGCCCTAATATTTTGGACATGCTGCTTCTAGTAGATTCAGG 371

DB 302 TTTCAGGCGCTGATGCCCTAATATTTTGGACATGCTGCTTCTAGTAGATTCAGG 361  
 QY 372 TCAATATGTATGATGATGCCACAGTTCATGAGCAAGCACTAGTCCCAACTACAGTTC 431  
 DB 362 TCAATATGTATGATGATGCCACAGTTCATGAGCAAGCACTAGTCCCAACTACAGTTC 421  
 QY 432 TTGTTGTAGCAGTGTGAACCTGTAAGTTGATGGAACAACAACATTTCTTCAACACGA 491  
 DB 422 TTGTTGTAGCAGTGTGAACCTGTAAGTTGATGGAACAACAACATTTCTTCAACACGA 481  
 QY 492 ACTTCCTGCTGACCTGCTAGTCCACCAACATGCTGTGGAAGATTTGCAAGTAT 551  
 DB 482 ACTTCCTGCTGACCTGCTAGTCCACCAACATGCTGTGGAAGATTTGCAAGTAT 541  
 QY 552 GCTTCGCTTTTCAAGATTGCTAGTATGTAAGGGGCAAAAGTCCATTCATTTGCTC 611  
 DB 542 GCTTCGCTTTTCAAGATTGCTAGTATGTAAGGGGCAAAAGTCCATTCATTTGCTC 601  
 QY 612 CATTAAGTCCACCAATGAAATTTATGCAATTTATTTGATTTGATGAGACACTATATAT 671  
 DB 602 CATTAAGTCCACCAATGAAATTTATGCAATTTATTTGATTTGATGAGACACTATATAT 661  
 QY 672 GTGCTGAACCTAATTTCTTATATTTTCTATTCCTGACACACTTTCTAGACAGCTG 731  
 DB 662 GTGCTGAACCTAATTTCTTATATTTTCTATTCCTGACACACTTTCTAGACAGCTG 721  
 QY 732 CCAGTTTGGAGATGCGCCCTCAAGAGCTTTAAACTATTTTATACATGCTTATATAC 791  
 DB 722 CCAGTTTGGAGATGCGCCCTCAAGAGCTTTAAACTATTTTATACATGCTTATATAC 781  
 QY 792 ATTCACCTATATG 803  
 DB 782 ATTCACCTATATG 793  
 RESULT 5  
 AF212223 881 bp mRNA linear HTC 22-MAY-2001  
 LOCUS Homo sapiens BM025 mRNA, complete cds.  
 DEFINITION AF212223  
 ACCESSION AF212223.1 GI:9437512  
 VERSION HTC.  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 881)  
 Zhang, Q.H., Ye, M., Wu, X.Y., Ren, S.X., Zhao, M., Zhao, C.J., Fu, G.,  
 Shen, Y., Fan, H.Y., Lu, G., Zhong, M., Xu, X.R., Han, Z.G., Zhang, J.W.,  
 Tao, J., Huang, Q.H., Zhou, J., Hu, G.X., Gu, J., Chen, S.J. and Chen, Z.  
 Cloning and functional analysis of cDNAs with open reading frames  
 for 300 previously undefined genes expressed in CD34+ hematopoietic  
 stem/progenitor cells  
 JOURNAL Genome Res. 10 (10), 1546-1560 (2000)  
 MEDLINE 20499367  
 PUBMED 11042152  
 REFERENCE 2 (bases 1 to 881)  
 AUTHORS Zhao, M., Song, H., Li, N., Peng, Y., Han, Z. and Chen, Z.  
 TITLE A novel gene expressed in human bone marrow  
 JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 881)  
 AUTHORS Zhao, M., Song, H., Li, N., Peng, Y., Han, Z. and Chen, Z.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-DEC-1999) Chinese National Human Genome Center at  
 Shanghai, 351 Guo. Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,  
 Shanghai 201203, People's Republic of China  
 FEATURES  
 source  
 1. 881  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="bone marrow"  
 214. 558  
 CDS

```

/codon_start-1
/product="BM025"
/protein_id="AF87325.1"
/db_xref="GI:9437513"
/translation="MDKRRRLRLYLIDKATLIMNGNAVSGDLANNFEPDLPSSERO
VAMLDQPYHEQATQSQITVLYVTSKTVKDFGNKHFFNONFLTLAQSTPNNNTWKIA
SDCFRQDMSSS"
BASE COUNT      258 a 170 c 177 g 276 t
ORIGIN
Query Match      67.68; Score 775; DB 11; Length 881;
Best Local Similarity 94.58; Pred. No. 2,6e-184;
Matches 855; Conservative 0; Mismatches 45; Indels 3; Gaps 2;

29 AATTTTGGCGTTGGCGGTTGCTGCTCTCATAGATGATGATCCGAGCCCT 88
   |||||
   1 AATTTTGGCGTTGGCGGTTGCTGCTCTCATAGATGATGATCCGAGCCCT 60
   |||||
   89 GCGGACCGGA-CAGTGAAGAGAGTGAAGCGGACGACGACGACGACGACG 147
   |||||
   61 GCGGACCGGAGCGCTGTTACAGTCCGCTTCTGATCCGGAAGAGCCGAGCTGGGA 120
   |||||
   148 GTCCCAAGTGGCAGC--TCTGTGATTTTAAACCTTATGTATGATGATGATGAT 205
   |||||
   121 ATAGCAGAGAGCGGCGGATGTTGGGATTTTAAACCTTATGTATGATGATGATGAT 180
   |||||
   206 GGTGAGAGAGTTTTCATATTTTCTATGAGACATGATTAAGAAACGCGGACCTAAC 265
   |||||
   181 GGTGAGAGAGTTTTCATATTTTCTATGAGACATGATTAAGAAACGCGGACCTAAC 240
   |||||
   266 AGGCTGATCTGAGACAAGCGGACCTTATATGGAATGGAATGCTGTTTCAAGGCTGAT 325
   |||||
   241 AGGCTGATCTGAGACAAGCGGACCTTATATGGAATGGAATGCTGTTTCAAGGCTGAT 300
   |||||
   326 GCCCTAAATATTTTGTGACATGCTCTTATGATGATGATGATGATGATGATGAT 385
   |||||
   301 GCCCTAAATATTTTGTGACATGCTCTTATGATGATGATGATGATGATGATGATGAT 360
   |||||
   386 TGGCAACCAATTCATGAGCAAGCAACGACGACGACGACGACGACGACGACGACGAC 445
   |||||
   361 TGGCAACCAATTCATGAGCAAGCAACGACGACGACGACGACGACGACGACGACGAC 420
   |||||
   446 GGAAGTGTGAAGTTGATGAGAAACAACATCTTCTCAACGAGACTCCCTGCTGACT 505
   |||||
   421 GGAAGTGTGAAGTTGATGAGAAACAACATCTTCTCAACGAGACTCCCTGCTGACT 480
   |||||
   506 GCTCAGTCCACTCCCAACAATTAAGTGTGAGAGATGAGAGTGTGCTCCGTTTCAA 565
   |||||
   481 GCTCAGTCCACTCCCAACAATTAAGTGTGAGAGATGAGAGTGTGCTCCGTTTCAA 540
   |||||
   566 GATTGCTGTAGTGAAGGGGCAAAAGTCCATTCATTTGCTCATTAAGTTCCAGCA 625
   |||||
   541 GATTGCTGTAGTGAAGGGGCAAAAGTCCATTCATTTGCTCATTAAGTTCCAGCA 600
   |||||
   626 ATGGAATTTATGATGATTTTGTGATGATGAGAGCACTATTAATATGCTGGAACATA 685
   |||||
   601 ATGGAATTTATGATGATTTTGTGATGATGAGAGCACTATTAATATGCTGGAACATA 660
   |||||
   686 TTTCTTAATATTTTCTATTCCTGTCGACGACCTTTTACAGAGTGGCAGTTGGAGCAT 745
   |||||
   661 TTTCTTAATATTTTCTATTCCTGTCGACGACCTTTTACAGAGTGGCAGTTGGAGCAT 720
   |||||
   746 TGGCCTTAAGAGCTTTAAACATATTTTGTGATGATGATGATGATGATGATGATGAT 805
   |||||
   721 TGGCCTTAAGAGCTTTAAACATATTTTGTGATGATGATGATGATGATGATGATGAT 780
   |||||
   806 ATTCCTTAATATTTATTAACACATGATCTGTGATGATGATGATGATGATGATGATGAT 865
   |||||
   781 ATTCCTTAATATTTATTAACACATGATCTGTGATGATGATGATGATGATGATGATGAT 840
   |||||
   866 CTATTCGAAAAATTAATCTTTTATATATTA 898
   |||||
   841 CTATTCGAAAAATTAATCTTTTATATATTA 873

```

```

RESULT 6
LOCUS   AUI23481
DEFINITION AUI23481 NT2RM2 Homo sapiens cDNA clone NT2RM200386 5', mRNA.
ACCESSION AUI23481
VERSION   AUI23481.1 GI:10948197
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 781)
AUTHORS  Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Salto,K., Yamamoto,J.,
           Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
           Isogai,T.
TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
           Salto,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
           Y., Sugano,S., Isogai,T.)
JOURNAL   Unpublished (2000)
COMMENT   Contact: Takao Isogai
           Genomics Laboratory
           Helix Research Institute
           1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
           Tel: 81-438-52-3975
           Fax: 81-438-52-3986
           Email: genomcs@hri.co.jp
           HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
           Research Institute; cDNA library construction; Department of
           Virology, Institute of Medical Science, University of Tokyo, and
           Helix Research Institute.
FEATURES
     source
       1..781
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone="NT2RM200386"
         /clone_lib="NT2RM2"
         /cell_type="teratocarcinoma"
         /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
         precursor cells"
BASE COUNT      233 a 129 c 138 g 278 t 3 others
ORIGIN
Query Match      62.18; Score 711.2; DB 9; Length 781;
Best Local Similarity 99.68; Pred. No. 2.9e-168;
Matches 713; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

431 CTGTGTGACAGTGTGAGTGTGATGAGAAACAACAACATTTCTTCAACCAG 490
   |||||
   1 CTGTGTGACAGTGTGAGTGTGATGAGAAACAACAACATTTCTTCAACCAG 60
   |||||
   491 AACTTCCTGCGATCGTCACTCCACCACTACTGTGTGGAAGATGCAAGTAT 550
   |||||
   61 AACTTCCTGCGATCGTCACTCCACCACTACTGTGTGGAAGATGCAAGTAT 120
   |||||
   551 TGCCTCGTTTCAAGATTTGGTCTAGTAAAGGGGCAAAAGTCCATTTCTCATTTGGT 610
   |||||
   121 TGCCTCGTTTCAAGATTTGGTCTAGTAAAGGGGCAAAAGTCCATTTCTCATTTGGT 180
   |||||
   611 CCATTAGTCCAGCAATTTGAATTTATGATGATGATGATGATGATGATGATGATGAT 670
   |||||
   181 CCATTAGTCCAGCAATTTGAATTTATGATGATGATGATGATGATGATGATGATGAT 240
   |||||
   671 TGTCTGAACCAATTTCTTATATTTTCTATTCCTGTCGACGACCTTTCTTACGAGCT 730
   |||||
   241 TGTCTGAACCAATTTCTTATATTTTCTATTCCTGTCGACGACCTTTCTTACGAGCT 300
   |||||
   731 GCCAGTTTGGAGCATTCGCCCTCTAGAGCTTTAAACATATTTTTCATGAGCTTATATA 790
   |||||
   301 GCCAGTTTGGAGCATTCGCCCTCTAGAGCTTTAAACATATTTTTCATGAGCTTATATA 360

```

QY 721 CATTCACATATGACATTTCTTATATATATATTAACACATGATCTTGGACTACATACAT 850  
 DB 361 CATTCACATATGACATTTCTTATATATATATTAACACATGATCTTGGACTACATACAT 420  
 QY 851 CACTGTGAAACCCAGCCCTATGCAAAATTAATCTTTTATATATATATCTATGGAATG 910  
 DB 421 CACTGTGAAACCCAGCCCTATGCAAAATTAATCTTTTATATATATATCTATGGAATG 480  
 QY 911 CAGCACAATATATACATCTGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 970  
 DB 481 CAGCACAATATATACATCTGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
 QY 971 GTAAACACAGTTGCTGTTTTCAGTTTAACTGATGATGATGATGATGATGATGATGATG 1030  
 DB 541 GTAAACACAGTTGCTGTTTTCAGTTTAACTGATGATGATGATGATGATGATGATGATG 600  
 QY 1031 TCAATTCAGTTCAGTGAATAATGTCACATTTAGTTTACATACATCTCTGACATCTG 1090  
 DB 601 TCAATTCAGTTCAGTGAATAATGTCACATTTAGTTTACATACATCTCTGACATCTG 660  
 QY 1091 GAATTCATATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1146  
 DB 661 GAATTCATATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 716  
 RESULT 7  
 LOCUS AU137118 770 bp mRNA linear EST 02-ANG-2002  
 DEFINITION AU137118 PLACE1 Homo sapiens cDNA clone IMAGE1005820 5', mRNA  
 ACCESSION AU137118  
 VERSION AU137118.1 GI:10997657  
 KEYWORDS EST  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 770)  
 Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,  
 Yamamoto, J., Nakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and  
 Isogai, T.  
 TITLE HRI human cDNA project  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3975  
 Fax: 81-438-52-3986  
 Email: genomics@hri.co.jp  
 HRI human cDNA project: 5'-63'-end one pass sequencing: Helix  
 Research Institute; cDNA library construction: Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 FEATURES  
 source  
 1. 770  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="PLACE1005820"  
 /clone\_lib="PLACE1"  
 /tissue\_type="placenta"  
 /note="Vector: PME18SFJ3"  
 BASE COUNT 237 a 157 c 139 g 232 t 5 others  
 ORIGIN  
 Query Match 61.6%; Score 706.2; DB 9; Length 770;  
 Best Local Similarity 97.1%; Pred. No. 5,3e-167;  
 Matches 748; Conservative 0; Mismatches 18; Indels 4; Gaps 3;  
 QY 130 AGAACAACACTGCTACAGAGTCCC-AGATGGCCACGCTCTGGAATTTAAACTTATGATG 188  
 DB 1 AGAACAACACTGCTACAGAGTCCCAGATGGCCACGCTCTGGAATTTAAACTTATGATG 60

QY 189 ATCAGCATGTAGACCTGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 248  
 DB 61 ATCAGCATGTAGACCTGCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120  
 QY 249 GAAAGCGGGCACTAACCAAGGCTGATCTGAGCAAAAGCCCACTTATGATGATGAAATG 308  
 DB 121 GAAAGCGGGCACTAACCAAGGCTGATCTGAGCAAAAGCCCACTTATGATGATGAAATG 180  
 QY 309 CTGTTTCAGGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 368  
 DB 181 CTGTTTCAGGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240  
 QY 369 AGGTCAATATGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 428  
 DB 241 AGGTCAATATGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
 QY 429 TTCTTGTGTGTGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 488  
 DB 301 TTCTTGTGTGTGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 360  
 QY 489 AGAATCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 548  
 DB 361 AGAATCTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
 QY 549 ATGCTTCCTGCTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 608  
 DB 421 ATGCTTCCTGCTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
 QY 609 GTCCATTAATTCAGCAATGTAATTTATGTAATTTATGTAATTTATGTAATTTATGTA 668  
 DB 481 GTCCATTAATTCAGCAATGTAATTTATGTAATTTATGTAATTTATGTAATTTATGTA 540  
 QY 669 TATGCTGCAAACTAATTTCTTATATATTTCTTATATTTCTTATATTTCTTATATTTCT 728  
 DB 541 TATGCTGCAAACTAATTTCTTATATTTCTTATATTTCTTATATTTCTTATATTTCT 600  
 QY 729 CTGCGAGTTTGAGACATGCTCCCTCAAGAGCTTTAAACTATTTTAACTATGCTTAA 788  
 DB 601 CTGCGAGTTTGAGACATGCTCCCTCAAGAGCTTTAAACTATTTTAACTATGCTTAA 660  
 QY 789 TACATTCCTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 847  
 DB 661 TACATTCCTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720  
 QY 848 ACTCAGCTGGAAGCC--AGCCTATGCAAAATTAATTTTATATAT 895  
 DB 721 ACTCAGCTGGAAGCCCAAGCCCAATTAATTTTATATATTTTAAAA 770

RESULT 8  
 LOCUS BE887056 745 bp mRNA linear EST 20-OCT-2000  
 DEFINITION BE887056 NIH-MGC\_71 Homo sapiens cDNA clone IMAGE:3908026 5',  
 mRNA sequence.  
 ACCESSION BE887056  
 VERSION BE887056.1 GI:10341960  
 KEYWORDS EST  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 745)  
 NIH-MGC <http://mhc.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@email.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM9719 row: h column: 11  
 High quality sequence stop: 722.

## FEATURES

Location/Qualifiers

1..745  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3908026"  
 /clone\_lib="NIH\_MGC\_71"  
 /tissue\_type="leiomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; NotI;  
 Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 2.1 kb."

## BASE COUNT

208 a 151 c 156 g 230 t

## ORIGIN

Query Match 59.6%; Score 682.8; DB 12; Length 745;  
 Best Local Similarity 98.0%; Pred. No. 4.1e-161;  
 Matches 720; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

46 CGGGTTGGCTCTTCATAGTATGATCATTCGCGACCCCTGCGACCGGACAGTGA 105  
 1 CGGGTTGGCTCTTCATAGTATGATCATTCGCGACCCCTGCGACCGGACAGTGA 60  
 106 GGAGGTAGGAGCGCGACACTGCCAGAACACACTGCTACAGGTCCCAATGGCCACGTC 165  
 61 GGAGGTAGGAGCGCGACACTGCCAGAACACACTGCTACAGGTCCCAATGGCCACGTC 120  
 166 TCTGATTTTAAACTATATGATCAGCATGTAGAGTGGTGGAGAGTTTGTCAATAT 225  
 121 TCTGATTTTAAACTATATGATCAGCATGTAGAGTGGTGGAGAGTTTGTCAATAT 180  
 226 TTACTATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 285  
 181 TTACTATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 286 CACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 345  
 241 CACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 346 CACATTTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 405  
 301 CACATTTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 406 AGCAACTCAGTCCCAAACTCAGTCTTGTGTGACCACTGGAAGTGAAGTGTGATG 465  
 361 AGCAACTCAGTCCCAAACTCAGTCTTGTGTGACCACTGGAAGTGAAGTGTGATG 420  
 466 AACAACAACAATTTCTTCAACCAAGATCTGCTGATGCTGATGCTGATGCTGATG 525  
 421 AACAACAACAATTTCTTCAACCAAGATCTGCTGATGCTGATGCTGATGCTGATG 480  
 526 TACTGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 585  
 481 TACTGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540  
 586 GGGCAAAAGTCCATTCATTTGGTCCATTAATTCAGCAATGGAATTTATGTAATTA 645  
 541 GGGCAAAAGTCCATTCATTTGGTCCATTAATTCAGCAATGGAATTTATGTAATTA 600  
 646 TTTGATGTAGAGCACTATATATGTCGTAATAATTTCTTAATTTATTTCTAT 705  
 601 TTTGATGTAGAGCACTATATATGTCGTAATAATTTCTTAATTTATTTCTAT 660  
 706 CCGTCAGACCTTTTCTAGAGCGTCCGAGTTGGAGCAATGCGCTTAAAGCTTTAA 765  
 661 CCGTCAGACCTTTTCTAGAGCGTCCGAGTTGGAGCAATGCGCTTAAAGCTTTAA 717  
 766 ACTATTTT 775  
 718 ATTTTAT 727

## RESULT 9

BI58206

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NIH-MGC <http://mgs.nci.nih.gov/>  
 1 (bases 1 to 782)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs.fremail.nih.gov  
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM1703 row: f column: 12  
 High quality sequence stop: 759.

## FEATURES

Location/Qualifiers

1..782  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5278475"  
 /clone\_lib="NIH\_MGC\_96"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"

/note="Organ: Brain; Vector: pBluescript (modified  
 pluescript KS+); Site: 1; BamHI; Site: 2; SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.3 kb and  
 normalized to 10<sup>5</sup>. This is a primary library enriched  
 for full-length clones and constructed using the  
 cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH-MGC Library."

## BASE COUNT

238 a 144 c 156 g 244 t

## ORIGIN

Query Match 59.5%; Score 681.4; DB 13; Length 782;  
 Best Local Similarity 98.0%; Pred. No. 9.2e-161;  
 Matches 722; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

169 GGATTTTAAACTATATGATCAGCATGATGATGATGATGATGATGATGATGAT 228  
 47 GGATTTTAAACTATATGATCAGCATGATGATGATGATGATGATGATGATGAT 106  
 229 CTATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 288  
 107 CTATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 166  
 289 CTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 348  
 167 CTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 226  
 349 ATTGCTTCTAGTACGATCCAGTCAATATGATGATGATGATGATGATGATGATG 408  
 227 ATTGCTTCTAGTACGATCCAGTCAATATGATGATGATGATGATGATGATGATG 286  
 409 AACTCAGTCCCAAACTCAGTCTTGTGTGACCACTGGAAGTGAAGTGTGATG 468  
 287 AACTCAGTCCCAAACTCAGTCTTGTGTGACCACTGGAAGTGAAGTGTGATG 346

Oy		469	CANAACAACATTTCTTCAACCAGAAACTCCGCGTCAGCTCAGTCCAGCCATGCCAACAATAC	528
Dd		347	CAAACACACATTCTTCTTCAACGAGAATCTTCGTGACTGCTCAGTCCATGCCAACAATAC	406
Oy		529	TGATGGGAAGATTGCAGAGTAGTATGCTCCGTTTTCAAGATTGGTCTAGTAGTTAAAGGG	588
Dd		407	TGATGGGAAGATTGCAGAGTAGTATGCTCCGTTTTCAAGATTGGTCTAGTAGTTAAAGGG	466
Oy		589	CAAAAAGTCCATTTCATTTGGTGTCATTAGTTCACGCAATTGAANTTTATGTGAATTATT	648
Dd		467	CAAAAAGCCATTTCATTTGGTGTCATTAGTTCACGCAATTGAANTTTATGTGAATTATT	526
Oy		649	TGATTTAGSAGAGCACTTAATATAGTGTGAAACAAATTTCTTAATATTTTCATTCT	708
Dd		527	GGATTGTAGSAGAGCACTTAATATAGTGTGAAACAAATTTCTTAATATTTTCATTCT	586
Oy		709	GTCAGCACCTTTCTTACAGCTGCCAGTTGGAGCATTTGCCCTTAAGAGCTTTAAACT	768
Dd		587	GTCAGCACCTTTCTTACAGCTGCCAGTTGGAGCATTTGCCCTTAAGAGCTTTAAACT	646
Oy		769	ATTTTTCACATGCC-TTATATACATTCCTCAATAATGACATTTCTTATATATATTAACA	827
Dd		647	ATTTTTCACATGCC-TTATATACATTCCTCAATAATGACATTTCTTATATATATTAACA	706
Oy		828	CATGATCTTGGTACTAACA--TACTCACTGTGAACCCAGCCTATTGCAAAATTAATCT	885
Dd		707	CATGATCTTGGTACTAACAATACCTGACCTGTGAACCAAGCCTATTGCAAAATTAATCT	765
Oy		886	TTTTATATATATATATCTGA 902	
Dd		766	TTTTATATATATATATCTGA 782	
RESULT 10				
LOCUS	AV1714243			
DEFINITION	AV1714243 DCB Homo sapiens cDNA clone DCBAS08 5', mRNA sequence.			
ACCESSION	AV1714243			
VERSION	AV1714243.1 GI:10795760			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE.	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	1 (bases 1 to 721)			
	Xu,X., Gu,Y., Lu,F., Qu,J., Zhao,M., Li,Y., Huang,Q., Zhou,J.,			
	Song,H., Gu,Y., Yang,Y., Gao,G., Xiao,H., Li,N., Qian,B., Gao,X.,			
	Cheng,Z., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu			
	,G., Cheng,Z. and Han,Z.			
	Homo sapiens cDNA DCB clones.			
	Unpublished (2000)			
TITLE	Contact: Zeguang Han			
JOURNAL	Chinese National Human Genome Center at Shanghai			
COMMENT	351 Gui Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai			
	201203, P. R. China			
	Tel.: 86-21-50801919(ex.45)			
	Fax: 86-21-50801922			
	Email: hanzg@chgc.sh.cn			
	This clone is available at CHGC in Shanghai.			
FEATURES	Location/Qualifiers			
source	1..721			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="DCBAS08"			
	/clone_lib="DCB"			
	/cell_type="dendritic cells"			
	/dev_stage="mature"			
	/lab_host="BM25.8"			
	/note="vector: pTriplex2; site_1: sfIla; site_2: sfIIB"			
BASE COUNT	199 a 146 c 170 g 206 t			
ORIGIN				

Query Match	58.2%	Score 666.8;	DB 10;	Length 721;
Best Local Similarity	99.68;	Pred. No. 4,3e-157;		
Matches 679;	Conservative 0;	Mismatches 2;	Indels 1;	Gaps 1

QY	1	TACTCCGGGGAGAAATGGGAGGGGTGGAAATTTTGTGCGTTTGGCGGGTTCCTCTCT	60
Db	9	TACTCCGGGGAGAAATGGGAGGGGTGGAAATTTTGTGCGTTTGGCGGGTTCCTCTCT	68
QY	61	TCATAGTATTGATCATCTCCGACGCCCTCGCGAGCCGACACGTGAGAGGTAGTGCACC	120
Db	69	TCATAGTATTGATCATCTCCGACGCCCTCGCGAGCCGACACGTGAGAGGTAGTGCACC	128
QY	121	GACACTGGCAGAAACACACTGCTTACAAGSTGCCAATGGCCACGTCGTGGAFTTTAAAC	180
Db	129	GACACTGGCAGAAACACACTGCTTACAAGSTGCCAATGGCCACGTCGTGGAFTTTAAAC	188
QY	181	TTATGTAGATAGGCAATGATAGAGCGTCGAGAGAGTTTGCATATATTCTATGACACAT	240
Db	189	TTATGTAGATAGGCAATGATAGAGCGTCGAGAGAGTTTGCATATATTCTATGACACAT	248
QY	241	GGATAAAGAGACGGGCACTAACCGAGCTGATCTGACAGAGCCACCTTAATGGA	300
Db	249	GGATAAAGAGACGGGCACTAACCGAGCTGATCTGACAGAGCCACCTTAATGGA	308
QY	301	TGGAATGCTGTTTCAGGGCTGGATGCCCTAAATATTTTTTGCACATTGCCCTTTCAG	360
Db	309	TGGAATGCTGTTTCAGGGCTGGATGCCCTAAATATTTTTTGCACATTGCCCTTTCAG	368
QY	361	TGAGTTCAGGTCATATGTAGATTGGCAACCACTTATATGAGCAGAGCAACTCAGTCCCA	420
Db	369	TGAGTTCAGGTCATATGTAGATTGGCAACCACTTATATGAGCAGAGCAACTCAGTCCCA	428
QY	421	AACTCACGTTCTGTTGTGACACAGTGAACCTGTGAAGTTTGAATGGAACAACAACATTT	480
Db	429	AACTCACGTTCTGTTGTGACACAGTGAACCTGTGAAGTTTGAATGGAACAACAACATTT	488
QY	481	CTTCACAGAACTTCTGCTGCTGACTGCTCAGTCCACTCCCAACATTACTGTGTGGAAGT	540
Db	489	CTTCACAGAACTTCTGCTGCTGACTGCTCAGTCCACTCCCAACATTACTGTGTGGAAGT	548
QY	541	TGCAAGTATGCTCTCCGTTTCAAGATGTGCTGTAAGTTAAAGGGGCAAAAGTCCATT	600
Db	549	TGCAAGTATGCTCTCCGTTTCAAGATGTGCTGTAAGTTAAAGGGGCAAAAGTCCATT	608
QY	601	CTCATTTGGTCATAGTTCACAGCAATGAAATTAATGGAATTAATTTGATTGTGGAAG	660
Db	609	CTCATTTGGTCATAGTTCACAGCAATTAATTTGATGGAATTAATTTGATTGTGGAAG	667
QY	661	CACCTAATATGCTGAGAACT 682	
Db	668	CACCTAATATGCTGAGAACT 688	

RESULT 11					
LOCUS	BT161278	743 bp	mRNA	linear	EST 25-SEP-2001
DEFINITION	603044031E1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5184559 5',				
ACCESSION	mRNA sequence.				
VERSION	BT161278				
KEYWORDS	BT161278.1	GI:15752856			
SOURCE	Est.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo				
JOURNAL	1 (bases 1 to 743)				
COMMENT	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
	National Institutes of Health, Mammalian Gene Collection (MGC)				
	unpublished (1998)				
	Contact: Robert Strausberg, Ph.D.				
	Email: <a href="mailto:cgabs@remail.nih.gov">cgabs@remail.nih.gov</a>				
	Tissue Procurement: Life Technologies, Inc.				
	cDNA Library Preparation: Life Technologies, Inc.				







|||||  
 Db 228 ATGGCTCTAGTGGATCCAGGTCATATGTTAATGTCACCAAGTCATGACCAAGC 287  
 OY 409 AACGAGGCCAACTACAGTCTTGTGTGACAGGAGTGAAGTTGATGAGAA 468  
 Db 288 AACGAGGCCAACTACAGTCTTGTGTGACAGGAGTGAAGTTGATGAGAA 347  
 OY 469 CAAACAACTTCTTCAACAGCAACTTCTGTGACAGTGAAGTTGATGAGAA 528  
 Db 348 CAAACAACTTCTTCAACAGCAACTTCTGTGACAGTGAAGTTGATGAGAA 407  
 OY 529 TGTGTGAGAGTTCAGAGTGTGCTCCGTTTTCAGAGTGTGATGAGAA 588  
 Db 408 TGTGTGAGAGTTCAGAGTGTGCTCCGTTTTCAGAGTGTGATGAGAA 467  
 OY 589 CAAACAACTTCTTCAACAGCAACTTCTGTGACAGTGAAGTTGATGAGAA 648  
 Db 468 CAAACAACTTCTTCAACAGCAACTTCTGTGACAGTGAAGTTGATGAGAA 527  
 OY 649 TGTGTGAGAGTTCAGAGTGTGCTCCGTTTTCAGAGTGTGATGAGAA 708  
 Db 528 TGTGTGAGAGTTCAGAGTGTGCTCCGTTTTCAGAGTGTGATGAGAA 587  
 OY 709 GTGAGACCTTTTTCAGAGTGTGCTCCGTTTTCAGAGTGTGATGAGAA 766  
 Db 588 GTGAGACCTTTTTCAGAGTGTGCTCCGTTTTCAGAGTGTGATGAGAA 647  
 OY 767 GTGAGACCTTTTTCAGAGTGTGCTCCGTTTTCAGAGTGTGATGAGAA 824  
 Db 648 GTGAGACCTTTTTCAGAGTGTGCTCCGTTTTCAGAGTGTGATGAGAA 707  
 OY 825 A-CACATGATCTTGTGA-CTACATGCTCA-CTGTGAGCCAGCTTTCAGAGTGA 881  
 Db 708 ACCACATGATCTTGTGA-CTACATGCTCA-CTGTGAGCCAGCTTTCAGAGTGA 767  
 OY 882 A--TCTTTTAAATATATATATATATATATATATATATATATATATATAT 930  
 Db 768 AATCTTAAAT 819  
 RESULT 13  
 Bg035698 878 bp mRNA linear EST 24-JAN-2001  
 LOCUS 60325770F1 NIH\_MGC\_90 Homo sapiens cDNA clone IMAGE:4413857 5'  
 DEFINITION mRNA sequence.  
 ACCESSION Bg035698  
 VERSION Bg035698.1 GI:12430092  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 878)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgep@briemall.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L14M10140 row: 1 column: 18  
 High quality sequence stop: 684.  
 Location/Qualifiers  
 1. 878  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4413857"  
 /clone\_1b="NIH\_MGC\_90"

/tissue\_type="adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: liver; Vector: pcMV-Sp08T6; Site: 1; NotI;  
 Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.7 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH-MGC Library."  
 BASE COUNT 259 a 145 c 167 g 307 t  
 ORIGIN  
 Query Match 53.4%; Score 612.2; DB 12; Length 878;  
 Best Local Similarity 95.5%; Pred. No. 2.4e-143;  
 Matches 696; Conservative 0; Mismatches 23; Indels 10; Gaps 6;  
 OY 427 AGTCTCTGTGAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 486  
 Db 1 AGTCTCTGTGAGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 60  
 OY 487 CCAGAACTTCTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 546  
 Db 61 CCAGAACTTCTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
 OY 547 TGAATGCTTCCGTTTTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 606  
 Db 121 TGAATGCTTCCGTTTTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
 OY 607 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666  
 Db 181 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 OY 667 AATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726  
 Db 241 AATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
 OY 727 AGTGCAGTGTGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786  
 Db 301 AGTGCAGTGTGAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360  
 OY 787 TATACATTCACATGATGATGATGATGATGATGATGATGATGATGATGATG 846  
 Db 361 TATACATTCACATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
 OY 847 TACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 906  
 Db 421 TACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 OY 907 ATGTGAGCACAAT 966  
 Db 481 ATGTGAGCACAAT 539  
 OY 967 TCTAGTAAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1024  
 Db 540 TCTAGTAAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 599  
 OY 1025 GCTTTTCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1078  
 Db 600 GCTTTTCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659  
 OY 1079 TCTGACATAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 1137  
 Db 660 TGGACCATAGTGAATGATGATGATGATGATGATGATGATGATGATGATG 719  
 OY 1138 TTTTGTGTTT 1146  
 Db 720 TTTTGTGTTT 728  
 RESULT 14  
 Bf087075 629 bp mRNA linear EST 19-OCT-2000  
 LOCUS Bf087075  
 DEFINITION PM2-H10342-130900-002-h05 H10342 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION Bf087075  
 VERSION Bf087075.1 GI:10892785

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 629) Dias Neto, E., Garcia Correa, R., Veljovski-Almeida, S., Britones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612-PM2-HT0342-130) 900-002-H056E3-2000-09-13644-1) Seq primer: puc 18 forward High quality sequence start: 3 High quality sequence stop: 611. Location/Qualifiers 1..629 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_1lb="HT0342" /dev_stage="Adult" /note="Organ: head,neck; Vector: puc18; Site_1: Smal; Site_2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT	187 a 129 c 131 g 182 t
ORIGIN	
Query Match	53.3%, Score 611, DB 12, Length 629;
Best Local Similarity	98.4%, Pred. No. 4,6e-143;
Matches 617; Conservative	0; Mismatches 10; Indels 0; Gaps 0;
OY	119 CCGACACTGCCAGAAACACACTGCTACAAAGGCCCAAGTGGCAGCTCTCGATTTTAAA 178
Db	2 CCGACACTGCCAGAAACACACTGCTACAAAGGCCCAAGTGGCAGCTCTCGATTTTAAA 61
OY	179 ACTATATAGATTCAGGCATGTAGAGCTGCTGAGAGTTTGTCATATTACTATAGACAA 238
Db	62 ACTATATAGATTCAGGCATGTAGAGCTGCTGAGAGTTTGTCATATTACTATAGACAA 121
OY	239 ATGATTAAGAAAGAGACGGGCACCTAACAGGCTGTATCTGGACAAAGCCACCTTAATATGG 298
Db	122 ATGATTAAGAAAGAGACGGGCACCTAACAGGCTGTATCTGGACAAAGCCACCTTAATATGG 181
OY	299 AATGAAATGCTGTTTACAGGGCTGATGCCCTTAATATTTTGTGACATATGCCCTTC 358
Db	182 AATGAAATGCTGTTTACAGGGCTGATGCCCTTAATATTTTGTGACATATGCCCTTC 241
OY	359 AGTAGATTCAGGTCATTAATGTAGATTGCCAACCAAGTCATGAGCAAGCAAGCACTAGTCC 418
Db	242 AGTAGATTCAGGTCATTAATGTAGATTGCCAACCAAGTCATGAGCAAGCAAGCACTAGTCC 301
OY	419 CAAACTCACTTCCTGTTGTGACACAGGAAAGCTGAAGTTTGATGGAACAAACAAAT 478

DB	FEATURES	source
Db	302 CAACCTACAGTTCTGTGTTGATACACAGTGAAGTGAAGTTGGATGCAAAACAACAT	361
QY	479 TTCTTCAACCCAGAACTTCTCTGCTGACTGCTCAGTCCACTCCCAACAATACTGTGTGAAG	538
Db	362 TTCTTCAACCCAGAACTTCTCTGCTGACTGCTCAGTCCACTCCCAACAATACTGTGTGAAG	421
QY	539 ATTCGAATGATGTTGGTCTCGTTTCCAAATGAGTGTACTACTATTAAAGGGCAAAAGTCCA	598
Db	422 ATTCGAATGATGTTGGTCTCGTTTCCAAATGAGTGTACTACTATTAAAGGGCAAAAGTCCA	481
QY	599 TTCTCATTTGGTCCATTTAGTTCACAGCAATTAATTAATTTGTAATTAATTTGATGTGTGA	658
Db	482 TGCATATTGGTCCATTCAGTCCAGCAATTAATTAATTTGTAATTAATTTGATGTGTGA	541
QY	659 AGCACTATAATATGCTGTGAACATAAATTTCTTAATATTTTCTAATTCCTGTACGACACT	718
Db	542 AGCACTATAATATGCTGTGAACATAAATTTCTTAATATTTAATTAACATATTCCTGTACGACACT	601
QY	719 TTCTAGCAGCTGCCAGTTTGAGCAT	745
Db	602 TTCTAGCAGCTGCCAGTTTGAGCAT	628
RESULT 15		
LOCUS	BC995586	629 bp mRNA linear EST 13-JUN-2001
DEFINITION	PM-1H0342-190201-007-h05 HT0342 Homo sapiens cDNA, mRNA sequence.	
ACCESSION	BC995586	
VERSION	BC995586.1 GI:14399656	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 629) Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Bittencourt, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Grindman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brustein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.	
COMMENT	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 20202653 Contact: Simpson, A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research. Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-27049922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tbl=PM2&tbl=PM2-HT0342-190201-007-h05&tbl=2001-02-19&tbl=1) Seq primer: puc 18 forward High quality sequence start: 3 High quality sequence stop: 611. Location/Qualifiers 1. 629 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="HT0342" /dev_stage="Adult" /note="Organ: head,neck; Vector: puc18; Site:1: Smal; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."	

BASE COUNT	187 a	129 c	131 g	182 t
ORIGIN				

Query Match	53.38;	Score 611;	DB 13;	Length 629;
Best Local Similarity	98.48;	Pred. No. 4.6e-143;		
Matches 617; Conservative	0;	Mismatches 10;	Indels 0;	Gaps 0;

QY	119	CCGCACTGGCAGAAACACACACTGCTACAAAGTCTCCAGATGGGCAAGCCTCTCTGGATTTTAA	178
Db	2	CCGCACTGGCAGAAACACACTGCTACAAAGTCTCCAGATGGGCAAGCCTCTCTGGATTTTAA	61
QY	179	ACTATGTAGATCAGGCATGTAGAGCTGCTGAGGAGTTGTCTAATATTTACTATGAGACA	238
Db	62	ACTATGTAGATCAGGCATGTAGAGCTGCTGAGGAGTTGTCTAATATTTACTATGAGACA	121
QY	239	ATGATATAAAGAAACAGGGGCACTAACACAGGCTGTATCTGAGCAAGGCCACCTTAATATGG	298
Db	122	ATGATATAAAGAAACAGGGGCACTAACACAGGCTGTATCTGAGCAAGGCCACCTTAATATGG	181
QY	299	AATGAAATGCTGTTTCAGGGCTGGATGCCCTAAATATTTTTTTGACACATTCCTTCT	358
Db	182	AATGAAATGCTGTTTCAGGGCTGGATGCCCTAAATATTTTTTTGACACATTCCTTCT	241
QY	359	AGTAGTCCAGGTCATATATTTAGATGGCAACACAGTTCATGACACAGCACTAGTCC	418
Db	242	AGTAGTCCAGGTCATATATTTAGATGGCAACACAGTTCATGACACACACTAGTCC	301
QY	419	CAAACTACAGTCTGTGTGTGACCAAGTGAACACTGGAATTTGAGAAACAAACAACT	478
Db	302	CAAACTACAGTCTGTGTGTGACCAAGTGAACACTGGAATTTGAGAAACAAACAACT	361
QY	479	TTCTTCAACCAAGAACTTCTGCTGACTGCTCAGTCCACTCCCAACATATCTGTGGAAG	538
Db	362	TTCTTCAACCAAGAACTTCTGCTGACTGCTCAGTCCACTCCCAACATATCTGTGGAAG	421
QY	539	ATTGCAATGATTTGGTTCCTCGTTTCAAGATTTGCTAGTAGTTAAAGGGGCAAAAGTCCA	598
Db	422	ATTGCAATGATTTGGTTCCTCGTTTCAAGATTTGCTAGTAGTTAAAGGGGCAAAAGTCCA	481
QY	599	TTTCATTTGGTCCATTAGTCCACCAATTTGAATTTATGTAAATTTATTTGATTGTAGA	658
Db	482	TGCTCATTTGGTCCATTAGTCCACCAATTTGAATTTATGTAAATTTATTTGATTGTAGA	541
QY	659	AGCACTATTAATATGTGCTGAACCTAAATTTCTTAATATTTTCTATTTCTGTGACAGCCT	718
Db	542	AGCACTATTAATATGTGCTGAACCTAAATTTCTTAATATTTTCTATTTCTGTGACAGCCT	601
QY	719	TTTTCAGACGCTGCAGTTGGAGAT 745	
Db	602	TTTTCAGACGCTGCAGTTGGAGAT 628	

```
Search completed: June 28, 2003, 03:34:25
Job time : 1928.73 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:27:44 ; Search time 62.6356 Seconds

(without alignments)  
5611.049 Million cell updates/sec

Title: US-09-763-902b-15

Perfect score: 1146  
1 taccctcgagagagatggtg.....tttaagtaatttttttt 1146

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 443362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/PCBUS.COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	40	3.5	2692	1 US-08-036-210-14	Sequence 14, Appl
C 2	40	3.5	2692	2 US-08-449-609-14	Sequence 14, Appl
C 3	39.4	3.4	1511	1 US-07-991-867B-8	Sequence 8, Appl
C 4	39.4	3.4	1511	1 US-08-107-755A-8	Sequence 8, Appl
C 5	39.4	3.4	1511	2 US-08-544-332-8	Sequence 8, Appl
C 6	39.4	3.4	1511	4 US-09-370-861A-8	Sequence 8, Appl
C 7	37	3.2	45360	4 US-09-453-702B-206	Sequence 206, App
C 8	37	3.2	45325	4 US-08-487-826B-13	Sequence 261, App
C 9	36.8	3.2	19124	2 US-08-441-638-651	Sequence 13, Appl
C 10	36.4	3.2	20674	4 US-09-641-638-651	Sequence 651, App
C 11	36.2	3.2	998	4 US-09-122-400B-5	Sequence 5, Appl
C 12	36	3.1	740	4 US-08-998-416-971	Sequence 971, App
C 13	35.8	3.1	9636	4 US-08-323-170B-1	Sequence 1, Appl
C 14	35.8	3.1	9636	4 US-08-954-441-1	Sequence 1, Appl
C 15	35.6	3.1	2251	4 US-08-991-677-11	Sequence 11, Appl
C 16	35.4	3.1	2030	2 US-08-705-937-7	Sequence 7, Appl
C 17	35.2	3.1	7218	1 US-08-232-463-14	Sequence 14, Appl
C 18	34.6	3.0	460	4 US-09-643-597-205	Sequence 205, Appl
C 19	34.6	3.0	1056	4 US-09-134-001C-1550	Sequence 1550, Ap
C 20	34.2	3.0	2949	4 US-09-412-554A-3	Sequence 3, Appl
C 21	34	3.0	998	4 US-09-122-400B-5	Sequence 5, Appl
C 22	34	3.0	1890	6 5312912-3	Sequence 5, Appl
C 23	33.8	2.9	7577	4 US-08-961-527-46	Sequence 46, Appl
C 24	33.8	2.9	90050	4 US-09-245-041-5	Sequence 5, Appl
C 25	33.6	2.9	2308	4 US-08-480-229C-28	Sequence 28, Appl
C 26	33.6	2.9	2308	4 US-08-659-235C-28	Sequence 28, Appl
C 27	33.6	2.9	3846	2 US-08-845-161A-5	Sequence 5, Appl

28	33.6	2.9	3846	4 US-09-270-751-5	Sequence 5, Appl
29	33.6	2.9	3846	4 US-09-168-218B-3	Sequence 3, Appl
C 30	33.6	2.9	4967	4 US-09-177-325-12	Sequence 12, Appl
C 31	33.6	2.9	4967	4 US-09-411-812A-12	Sequence 12, Appl
C 32	33.6	2.9	4967	4 US-09-590-113-12	Sequence 12, Appl
C 33	33.6	2.9	51952	3 US-08-947-823-1	Sequence 1, Appl
C 34	33.4	2.9	1004	4 US-08-714-918-69	Sequence 69, Appl
C 35	33.4	2.9	1004	4 US-09-265-315-69	Sequence 69, Appl
C 36	33.4	2.9	1004	4 US-09-265-315-69	Sequence 69, Appl
C 37	33.4	2.9	1004	4 US-09-266-417-69	Sequence 69, Appl
C 38	33.4	2.9	1029	4 US-09-134-001C-1878	Sequence 1878, Ap
C 39	33.4	2.9	18443	4 US-09-078-294-6	Sequence 6, Appl
C 40	33.4	2.9	169998	4 US-09-676-610B-24	Sequence 24, Appl
C 41	33.2	2.9	1847	4 US-09-381-849-5	Sequence 5, Appl
C 42	33.2	2.9	36159	4 US-09-749-588-3	Sequence 3, Appl
C 43	33	2.9	720	4 US-08-513-014A-4	Sequence 4, Appl
C 44	33	2.9	810	4 US-09-072-993C-5	Sequence 5, Appl
C 45	33	2.9	1353	4 US-08-913-014A-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1  
US-08-036-210-14/C  
Sequence 14, Application US/08036210  
Patent No. 5585233  
GENERAL INFORMATION:  
APPLICANT: Moller, Niels P.H.  
APPLICANT: Moller, Karin B.  
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
PHOSPHATASE  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/036,210  
FILING DATE: 23-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2692 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 2149  
OTHER INFORMATION: /note="N=x-unknown nucleotide"  
US-08-036-210-14  
Query Match 3.5%; Score 40; DB 1; Length 2692;  
Best Local Similarity 53.9%; Pred. No. 0.09;



Best Local Similarity 49.3%; Pred. No. 0.1;  
Matches 103; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 768 TATTTTTCACGCTTATATACATTCACATATACATTTCTTAAATATTAACA 827  
|| ||| | |||| | | | | |||| |  
Db 1013 TAAATATTTAT 954  
QY 828 CATGATCTGTGACTACATCTACTGACGCCATTTGCAAAAATTAATCTTT 887  
|| ||| | |||| | | | | |||| |  
Db 953 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 894  
TAT  
QY 888 TAT 947  
|| ||| | |||| | | | | |||| |  
Db 893 TAAATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 834  
QY 948 TTGTTATTAGGTTAATTTCTAGTAAA 976  
|| ||| | |||| | | | | |||| |  
Db 833 TTTTCATGATTAATTTTGTGAAAA 805

## RESULT 4

US-08-107-755A-8/C  
; Sequence 8, Application US/08107755A  
; Patent No. 5721352  
; GENERAL INFORMATION:  
; APPLICANT: Moyer, Richard W.  
; APPLICANT: Hall, Richard L.  
; APPLICANT: Gruidl, Michael E.  
; TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: David R. Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: U.S.A.  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/107/755A  
; FILING DATE: 19-AUG-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/827,658  
; FILING DATE: 30-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/657,584  
; FILING DATE: 19-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (904) 375-8100  
; TELEFAX: (904) 372-5800  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1511 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Amsacta moorei entemopoxvirus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: Complement (18..218)  
; FEATURE:  
; NAME/KEY: CDS

LOCATION: complement (234..782)

FEATURE: CDS  
NAME/KEY: 852..1511  
LOCATION: 852..1511

## US-08-107-755A-8

Query Match 3.4%; Score 39.4; DB 1; Length 1511;  
Best Local Similarity 49.3%; Pred. No. 0.1;  
Matches 103; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 768 TATTTTTCACGCTTATATACATTCACATATACATTTCTTAAATATTAACA 827  
|| ||| | |||| | | | | |||| |  
Db 1013 TAAATATTTAT 954  
QY 828 CATGATCTGTGACTACATCTACTGACGCCATTTGCAAAAATTAATCTTT 887  
|| ||| | |||| | | | | |||| |  
Db 953 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 894  
TAT  
QY 888 TAT 947  
|| ||| | |||| | | | | |||| |  
Db 893 TAAATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 834  
QY 948 TTGTTATTAGGTTAATTTCTAGTAAA 976  
|| ||| | |||| | | | | |||| |  
Db 833 TTTTCATGATTAATTTTGTGAAAA 805

## RESULT 5

US-08-544-332-8/C  
; Sequence 8, Application US/08544332  
; Patent No. 5935777  
; GENERAL INFORMATION:  
; APPLICANT: Moyer, Richard W.  
; APPLICANT: Hall, Richard L.  
; APPLICANT: Gruidl, Michael E.  
; TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System  
; NUMBER OF SEQUENCES: 77  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Gerard H. Bencen  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: FL  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/544,332  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/991,867  
; FILING DATE: 07-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/107,755  
; FILING DATE: 19-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO 92/14818  
; FILING DATE: 12-FEB-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/827,685  
; FILING DATE: 30-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/657,584  
; FILING DATE: 19-FEB-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bencen, Gerard H.  
; REGISTRATION NUMBER: 35,746  
; REFERENCE/DOCKET NUMBER: UF114.C4  
; TELECOMMUNICATION INFORMATION:

COUNTRY: US  
 ZIP: 53701-2113  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Word Perfect 8.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/453,702B  
 FILING DATE: 03-Dec-1999  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/110,955  
 FILING DATE: 04-DEC-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seay, Nicholas J.  
 REGISTRATION NUMBER: 27386  
 REFERENCE/DOCKET NUMBER: 960296.95017  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (608) 251-5000  
 TELEFAX: (608) 251-9166  
 INFORMATION FOR SEQ ID NO: 206:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 43360  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 SEQUENCE DESCRIPTION: SEQ ID NO: 206:  
 US-09-453-702B-206

Query Match 3 2% Score 37; DB 4; Length 43360  
 Best Local Similarity 46.98% Pred No. 2.5; 130; Indels  
 Matches 115; Conservative 0; Mismatches 130; Indels

OY	856	TGAACCCAGCCTATTGCAAAAATAATCTTTTAAATATPTATTCATGGAGATCGACA	915
Db	22674	TGTACAGACTAACAAAGAAGATTAATCCCTCTTGACCTAAGATCCCCCTCACGGAGATTA	22733
OY	916	CATATTAAACACTGTGGSAAGAGTAGAGTTTTTGGTTATTAGGTAAATTTTCTAGTAA	975
Db	22734	CGAGGAACACAGGTTTGAAAGAAAGGAGGATTAATATGTCCTCATTTAATTTCTTTGCCAA	22793
OY	976	ACAGCTTGCCGTGTTTTGAGTTAAACACGCAGTAATGCCATTTAATATATGGCTTTTCAAA	1035
Db	22794	GCCTTATCTACACC GCCGCAAAAAACCGCTTAATATAGCGTTAGAAACATCCATTTTTCGGT	22853
OY	1036	TCAGTTCAAGTAAATATAGTACATTTAGGTTTACATTACTACTGACATACCTGGAATT	1095
Db	22854	GTTATTTCTTTGTAAGTTTGTACACATATATCATCTTTTGGGCAATCTTTAATAAAGTAAACA	22913
OY	1096	GCATA 1100	
Db	22914	GCATA 22918	

## RESULT 8

Sequence 61. Application US/09453702B  
Patent No. 6365723  
GENERAL INFORMATION:  
APPLICANT: Blattner, Frederick R.  
Burland, Valerie  
Perna, Nicole T.  
Plunkett, Guy  
Welch, Rod  
TITLE OF INVENTION: NO. 6365723el Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Plinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 1.44MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-Dec-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ ID NO: 261:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45325 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 261:  
US-09-453-702B-261

Query Match	3.28;	Score 37;	DB 4;	Length 45325;
Best Local Similarity	46.98;	Pred. NO. 2.5;		
Matches 115;	Conservative 0;	Mismatches 130;	Indels 0;	Gaps 0;

OY	856	TGAACCCAGCCTATTGCAAAAATAAATCTTTTAAATTAATTTATCATGGAGTCAGCA	915
Db	23576	TGTACAGAGACTAACAGAAGAATATACCCTTTCAAGCTAAGATCCCCTCCAGGGAGTTA	23635
OY	916	CAATATTAACACTCTGGAGAAAGTGCGATTTTTTGGTTATTAAGTTAAATTTCTGTAAAA	975
Db	23636	CGAGGAACAACGAGTTGCAAGAAATGGAGTATATATGCTGCATTTAAATTTCTGCCANA	23695
OY	976	ACAGCTTGCCGTTTTGAGTTAAACAGCACGCGTAAAGCCATTTTAAATATATAGCTTTTCAAA	1035
Db	23696	GCCCTATCTACACCCGCCAGAAAACAGCTTAATAAGCTTGAGAAACATCCATTTTCCGT	23755
OY	1036	TCAAGTTACGTAAATAATAGTACAGATTYTAGTTTACATACTACTGCAATACTGCAANT	1095
Db	23756	GTTATTTCTGTGTAAGCTTTGTACACATAATCATCTTTTGGGCGATCTTTAATAAGTGAACA	23815
OY	1096	GCATA 1100	
Db	23816	GCATA 23820	

## RESULT 9

US-08-487-826B-13/C  
Sequence 13, Application us/08487826B  
Patent No. 5993827  
GENERAL INFORMATION:  
APPLICANT: Sim, Kim L.  
APPLICANT: Chlitis, Chetan  
APPLICANT: Miller, Louis H.  
APPLICANT: Peterson, David S.  
APPLICANT: Su, Xin-zhaun  
APPLICANT: Wellens, Thomas E.  
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe Martens Olson & Bear  
STREET: 620 Newport Center Drive  
CITY: Newport Beach  
STATE: California  
COUNTRY: US  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,826B  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Israelson, Ned  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER: NIH121.001CPI  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
-type: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-487-826B-13

Query Match	3.2%	Score 36.8;	DB 2;	Length 19124;
Best Local Similarity	44.5%;	Pred. No. 1.9;		
Matches 146;	Conservative	0;	Mismatches 182;	Indels 0;
			Gaps	0;



```
QY 612 CATTAAGTCCAGCAATGGAATTTATGTAATTTGATTTGAGAACCACTATAATAT 671
|||||
DB 6310 CATTAATAATATATATATATCAATATATGATTTATATAAAGTTCATCTTATATAT 6251
QY 672 GTGCTGAACTAAATTTCTTATATATTTTCTGACAGACCTTTTTCAGACGCTG 731
|||||
DB 6250 TATATAAGAAATATTTGTTATATAATATACATATATGCTACTATATAATTTAA 6191
QY 732 CCAGTTTGAGCCCTTCCTCAGACGCTTAAACATATTTTTCATGACCTTATATAC 791
|||||
DB 6190 TATATCTTTAAAGTATATATATATATATATATATATATATATATATATATAT 6131
QY 792 ATCCCATATGATCTTATATATATATATATATATATATATATATATATATATAT 851
|||||
DB 6130 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 6071
QY 852 ACTGTACACCCAGCCCTATTCGCAAAATTAATCTTTTATATATATATATATAT 911
|||||
DB 6070 ATATGATAGATATATATATATATATATATATATATATATATATATATATAT 6011
QY 912 AGCACAATATACACTCTGGGAGAGT 939
DB 6010 ATATATACAAAGAAAGACAGACAGAGT 5983

RESULT 10
US-09-641-638-651
Sequence 651, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET.051CPI
CURRENT APPLICATION NUMBER: US/09/641,638
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1123..3123
OTHER INFORMATION: 5' regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3671..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5860
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5996..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon

LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc.feature
LOCATION: 17555..20674
OTHER INFORMATION: 3' regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
```



Db 744 CTTAANGAAGTGAATAATTTCTAATAATATTTTGTAAAAAAGTGAAGAAAAAGC 685

Qy 858 AACCCAGCTATTCGCAAAATAATCTTTTATATATATCTATGAGATGCGACCA 917

Db 684 AGAAGATTAATTTCTAAGATTAATAAACTGAAAAAAGCTTAATATTTTAACTAAA 625

Qy 918 ATATACACTCTGGGAGAACTGAGTTTGGTTATATAGTTAATTTCTAGTAAAC 977

Db 624 AACTGAAAAAAGAAAGAAATATTTTGTAGTTTTCAAAAAATATGTTTAA 565

Qy 978 ACGTGCCTGTTTCAGTTA 998

Db 564 ATATTTTCAGCTTTTAA 544

RESULT 12

US-08-998-416-971

Sequence 971, Application US/08998416

Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Pohlmann, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jurgen

APPLICANT: Knechtle, Philipp

APPLICANT: Redischung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHVA GOSYPPI

NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6239264artis Corporation

STREET: 3054 Cornwalls Road

CITY: Research Triangle Park

STATE: No. 6239264th Carolina

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416

FILING DATE: 24-DEC-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97

FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 971:

SEQUENCE CHARACTERISTICS:

LENGTH: 740 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: FAG1598RP

US-08-998-416-971

Query Match

Best Local Similarity 47.3%; Pred. No. 0.68;

Matches 142; Conservative 0; Mismatches 155; Indels 3; Gaps 1;

616 ACTTCAGCAATGTAATTTATGTAATTTATGTAAGCACTATATATGTC 675

Db 402 AGAACCTAATTTCAATCAATTAATTAATGATGATGTTGAGTGAATCAATACATTA 461

Qy 676 TGAAGTAAATTTCTTAATATTTTCTATTCCTGTCAGACCTTTCTAGACGTCGAG 735

Db 462 ATTAAGTAAATTTAAATTAATTAATTTGATTTCTATATGCAATATATTTAAAT 521

Qy 736 TTGGACATTTGCCCTTAAGAGCTTAAACT--ATTTTTCATGCTTAATCA 792

Db 522 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 581

Qy 793 TTCCATTAAGACATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTA 852

Db 582 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 641

Qy 853 CTGTGAACCCAGCTATTCGCAAAATAATCTTTTATATATATCTATGAGATGCA 912

Db 642 TAGGAATGAACTAATTAACCATTAAGATTAATTAATTAATTAATTAATTA 701

RESULT 13

US-08-323-1708-1/c

Sequence 1, Application US/083231708

Patent No. 5733772

GENERAL INFORMATION:

APPLICANT: Williamson, Kim C.

APPLICANT: Kaslow, David C.

TITLE OF INVENTION: Cloning and Expression of Plasmodium

TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf230

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/323,1708

FILING DATE: 13-OCT-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/010,409

FILING DATE: 29-JAN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Quine, Jonathan A.

REGISTRATION NUMBER: P-41,261

REFERENCE/DOCKET NUMBER: 015280-113100US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 9636 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 149..9556

US-08-323-1708-1

Query Match

Best Local Similarity 47.0%; Pred. No. 2.7;

Matches 147; Conservative 0; Mismatches 162; Indels 4; Gaps 1;

743 CATTCCTCTAAGAGCTTAATAACTATTTTTCATGATGCTTATATATCCACTAT 802

US-08-954-441-1

[illegible]

Mon Jun 30 08:51:38 2003

us-09-763-902b-15.rml

Page 10

Db 960 CCAATTGTATATAAACCAAAATCT 985

Search completed: June 28, 2003, 03:36:24  
Job time : 64.6356 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:39:51 ; Search time 295.199 Seconds  
(without alignments)  
8742.532 Million cell updates/sec

Title: US-09-763-902b-15

Perfect score: 1146

Sequence: 1 taccctccggagagaatggyg.....tttaagtaattttgtttt 1146

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1146	100.0	1146	21	AAA08041 Human protein tran
2	1136.4	99.2	2581	24	ABA05754 Human HNF2 associa
3	1082	94.4	2994	23	ABV74535 Human prostate exp
4	1064	92.8	2692	22	AAH17938 Human cDNA sequenc
5	978	85.3	1387	22	AA526214 Human cDNA encodin
6	747.2	65.2	8918	22	AA526708 Human genomic DNA
7	747.2	65.2	8919	22	AA526709 Human genomic DNA
8	491.2	29.0	716	22	AAH07014 Human cDNA clone (
9	332.4		1035	23	AA575617 DNA encoding novel

10	251.6	22.0	830	21	AAA08036 Human protein tran
11	251.6	22.0	864	22	AA526215 Human cDNA encodin
12	220	19.2	245	23	ABV07047 Human prostate exp
13	220	19.2	441	23	ABV36986 Human prostate exp
14	184.8	16.1	318	23	AA558525 cDNA #1201 encodin
15	60	5.2	60	24	ABN38056 Human spliced tran
16	55.6	4.9	805	23	ABN38056 Human spliced tran
17	49.2	4.3	17738	24	ABL16401 Drosophila melanog
18	47.2	4.1	13123	24	ABL16401 Drosophila melanog
19	47.2	4.1	13123	24	ABL16401 Drosophila melanog
20	46.6	4.1	2932	23	ABL16400 Human immune syste
21	46.6	4.1	4050	23	ABL09616 Human immune syste
22	46.6	4.1	7883	23	ABL08308 Human immune syste
23	45.6	4.0	6062	24	ABL34079 Human immune syste
24	45.6	4.0	15161	24	ABL70458 Human immune syste
25	45.6	4.0	15161	24	AA561423 Human gene regulat
26	45.2	3.9	33053	24	ABO67006 Human angiotensin
27	45	3.9	6002	24	ABL32952 Human immune syste
28	45	3.9	6365	24	ABL32124 Human immune syste
29	44.8	3.9	5559	24	ABL32338 Human immune syste
30	44.4	3.9	50849	24	ABN87883 Human glutathione
31	44	3.8	6831	24	ABL33487 Human immune syste
32	44	3.8	7047	24	ABK28386 DNA transcription
33	43.8	3.8	6167	24	ABK28370 DNA transcription
34	43.8	3.8	15732	22	AA545389 DNA transcription
35	43.8	3.8	15732	24	ABK28234 DNA transcription
36	43.6	3.8	9741	24	ABL33323 Human immune syste
37	43.6	3.8	47108	24	ABK31511 Signal transductio
38	43	3.8	5689	22	AA545384 Chemically pretrea
39	43	3.8	5689	22	AA546426 Tumour suppressor
40	43	3.8	5689	24	ABK28226 DNA transcription
41	42.8	3.7	11812	22	AA545501 Chemically pretrea
42	42.8	3.7	11812	22	AA546741 Tumour suppressor
43	42.8	3.7	11812	24	ABL34118 Human immune syste
44	42.8	3.7	11812	24	ABK28431 DNA transcription
45	42.8	3.7	12705	24	ABL32149 Human immune syste

#### ALIGNMENTS

##### RESULT 1

AAA08041 ID AAA08041 standard; cDNA; 1146 BP.

AC AAA08041;

DE 19-JUN-2000 (first entry)

DE Human protein transport molecule (PRM) encoding cDNA SEQ ID NO:15.

KW Human; protein transport molecule; PRM; diagnosis; cytostatic;

KW antidiabetic; antidiabetic; immunosuppressant; antidiabetic; osteoporotic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

KW antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;

PI Gorgone GA, Baughn MR, Patterson C;  
 XX WPI: 2000-256642/22.  
 DR P-PSDB; AAY82323.  
 XX New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis  
 XX  
 PS Claim 9, Page 74, 75pp: English.  
 CC AA08035 to AA08042 encode the human protein transport-associated  
 CC molecules (PRAMs) given in AAY82317 to AAY82324. The PRAMs have  
 CC cytosolic, antitubular, antileukemic, immunosuppressive,  
 CC antileukemic, antileukemic, antileukemic, antileukemic,  
 CC antileukemic, osteopathic, dermatological, antineoplastic, antiproliferative,  
 CC hepatocellular, antiapoptotic, antitubular and antileukemic activities, and  
 CC regulate protein transport. PRAM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PRAM. PRAM polynucleotides are useful for  
 CC diagnosing conditions associated with PRAM, comprising detecting PRAM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the  
 CC biological sample. Diseases prevented, treated or diagnosed include cell  
 CC proliferative disorders such as cancers, immune disorders, secretory  
 CC disorders and other conditions associated with abnormal vesicle  
 CC trafficking, such as allergies, asthma, urticaria and autoimmune  
 CC haemolytic anaemia. Anti-PRAM antibodies may be used as antagonists, as  
 CC a targeting or delivery mechanism for bringing pharmaceutical agents  
 CC into cells or tissues expressing PRAM and for diagnosis of PRAM-related  
 CC disorders. PRAM, its catalytic or immunogenic fragments are useful for  
 CC drug screening using libraries of compounds. PRAM polynucleotides are  
 CC useful for generating hybridisation probes useful in mapping the  
 CC naturally occurring genomic sequences.  
 CC  
 XX Sequence 1146 BP: 331 A; 212 C; 228 G; 375 T; 0 other:  
 SQ  
 Query Match 100.0%; Score 1146; DB 21; Length 1146;  
 Best Local Similarity 100.0%; Pred. No. 4,2e-291;  
 Matches 1146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 CTTCACCAAGAACTTCCTGCTGACGTCCACATCCCAACAAATCTGTCGAGAT 540  
 OY 541 TGCAGATGATGCTCCGCTTTTCAAGATGGTCTAGATGTAAGGGGCAAAAGTCATT 600  
 DB 541 TGCAGATGATGCTCCGCTTTTCAAGATGGTCTAGATGTAAGGGGCAAAAGTCATT 600  
 OY 601 CTGATTTGATGCTAGATGTAAGATGTAAGATGTAAGATGTAAGATGTAAGATGTAAG 660  
 DB 601 CTGATTTGATGCTAGATGTAAGATGTAAGATGTAAGATGTAAGATGTAAGATGTAAG 660  
 OY 661 CACTATATATGCTGCAAACTAAATTTCTTAATATTTCTATCTGTCAGACCTTT 720  
 DB 661 CACTATATATGCTGCAAACTAAATTTCTTAATATTTCTATCTGTCAGACCTTT 720  
 OY 721 TCTGACGCTGCGGCTTTGGAGCTTCCCTTAAGAGCTTAAAGATTTTTCAT 780  
 DB 721 TCTGACGCTGCGGCTTTGGAGCTTCCCTTAAGAGCTTAAAGATTTTTCAT 780  
 OY 781 GCCTTATATACATTCACATATGACATCTTATATATATATATTAACACATGATCTGGTA 840  
 DB 781 GCCTTATATACATTCACATATGACATCTTATATATATATATTAACACATGATCTGGTA 840  
 OY 841 CTACATATCTGCTGTAAGCCAGCCCTTATGCAAAATATATCTTTTATATATATATC 900  
 DB 841 CTACATATCTGCTGTAAGCCAGCCCTTATGCAAAATATATCTTTTATATATATATC 900  
 OY 901 TATGAGATGTCAGACATATATTAACATCTGGAGAGTGGATTTTGGTTATAGT 960  
 DB 901 TATGAGATGTCAGACATATATTAACATCTGGAGAGTGGATTTTGGTTATAGT 960  
 OY 961 TAATTTCTAGTAAACAGTGGCTGTTTCACTGTAACATGTAAGCCATTTTATA 1020  
 DB 961 TAATTTCTAGTAAACAGTGGCTGTTTCACTGTAACATGTAAGCCATTTTATA 1020  
 OY 1021 TATGCTTTTCAATCAGTTCAGTGAATATGATGATGATGATGATGATGATGATGATG 1080  
 DB 1021 TATGCTTTTCAATCAGTTCAGTGAATATGATGATGATGATGATGATGATGATGATG 1080  
 OY 1081 TGACATCTGGAATGCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
 DB 1081 TGACATCTGGAATGCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1140  
 OY 1141 TGTATT 1146  
 DB 1141 TGTATT 1146

RESULT 2  
 ABA05754  
 ID ABA05754 standard; cDNA; 2581 BP.  
 XX  
 AC ABA05754;  
 XX  
 DT 04-MAR-2002 (first entry)  
 XX  
 DE Human NTF2 associated protein 16 coding sequence.  
 XX  
 KM Human; NTF2 associated protein 16; cancer; HIV infection; gene therapy;  
 KM ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 147..575  
 FT FT /\*tag= a  
 XX FT /product= "NTF2 associated protein 16"  
 XX PN CNI33332-A.  
 XX PD 19-SEP-2001.  
 XX PF 15-MAR-2000; 2000CN-0114918.







oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH1628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2692 BP; 853 A; 471 C; 478 G; 890 T; 0 other.

Query Match 92.8%; Score 1064; DB 22; Length 2692;  
Best Local Similarity 100.0%; Pred. No. 2e-269;  
Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 83 AGCCCTGCGGACGCGACGAGAGTGTGACGCCGACACGCGACAGACACACTGCT 142
D 196 AGCCCTGCGGACGCGACGAGAGTGTGACGCCGACACGCGACAGACACACTGCT 255
QY 143 ACAAGTCCAGATGCCAGCTCTGTGATTTAAACTTATGTAGATCAGCATGTAGA 202
D 256 ACAAGTCCAGATGCCAGCTCTGTGATTTAAACTTATGTAGATCAGCATGTAGA 315
QY 203 GCTGCTGAGAGTGTGCAATATTACTATGAGACAATGATGATAAGAGAGGACACTA 262
D 316 GCTGCTGAGAGTGTGCAATATTACTATGAGACAATGATGATAAGAGAGGACACTA 375
QY 263 ACCAGGCTGATCTGGACAAGGCCACCTTAATGAAATGAAATGCTGTTACAGGGCTG 322
D 376 ACCAGGCTGATCTGGACAAGGCCACCTTAATGAAATGAAATGCTGTTACAGGGCTG 435
QY 323 GATGCCCTAATAATTTTTTTGACACATTTGCTGCTGAGTCCAGGCTCAATATGTTA 382
D 436 GATGCCCTAATAATTTTTTTGACACATTTGCTGCTGAGTCCAGGCTCAATATGTTA 495
QY 383 GATTGCCAACCGATTCATGAGCAAGCAACATGCTCCAAATACAGTTCTTGTGTGAC 442
D 496 GATTGCCAACCGATTCATGAGCAAGCAACATGCTCCAAATACAGTTCTTGTGTGAC 555
QY 443 AGTGAAGCTGTAAGTGTGATGAAACAACAACATTTCTTCAACCGAATCTCTGTG 502
D 556 AGTGAAGCTGTAAGTGTGATGAAACAACAACATTTCTTCAACCGAATCTCTGTG 615
QY 503 ACTGCTCAGTCACTCCCAACAATAGTGTGGAAGATGGAATGATGCTTCCGTTT 562
D 616 ACTGCTCAGTCACTCCCAACAATAGTGTGGAAGATGGAATGATGCTTCCGTTT 675
QY 563 CAAGATTGGTCTAGTAAAGGGGCAAAAGTCCATCTCATTTGGCTCAATAGTTCA 622
D 676 CAAGATTGGTCTAGTAAAGGGGCAAAAGTCCATCTCATTTGGCTCAATAGTTCA 735
QY 623 GCATTTAAATTTATGGAATTTATTTGATTTGAAAGCACTAATTAATGCTGAAACT 682
D 736 GCATTTAAATTTATGGAATTTATTTGATTTGAAAGCACTAATTAATGCTGAAACT 795
QY 683 AAATTTCTTAATTTTCTATTCCTGTCAGCACTTTTCTAGAGCTGCCAGTTGGAG 742
D 796 AAATTTCTTAATTTTCTATTCCTGTCAGCACTTTTCTAGAGCTGCCAGTTGGAG 855
QY 743 CATTTGCCCTTAAGAGCTTAAACATTTTCTTACATGCTTATATACATTCACATAT 802
D 856 CATTTGCCCTTAAGAGCTTAAACATTTTCTTACATGCTTATATACATTCACATAT 915
QY 803 GACATTTCTTAATTAATTAACACATGATCTTGTACTAATACATCTGTAACCC 862
D 916 GACATTTCTTAATTAATTAACACATGATCTTGTACTAATACATCTGTAACCC 975
QY 863 AGCCTATGCAAAATTAATCTTTTATATATATATCTATGAGATGCGCAAAATATA 922
```

```
DB 976 AGCCTATGCAAAATTAATCTTTTATATATATATCTATGAGATGCGCAAAATATA 1035
QY 923 AACTCTGGAGAAGAGTGGAGTATTTTGGTATTAAGTTAATTTCTAGTAAACACGTT 982
D 1036 AACTCTGGAGAAGAGTGGAGTATTTTGGTATTAAGTTAATTTCTAGTAAACACGTT 1095
QY 983 GCCGTGTTTCAAGTAAACCTGCTAATGCCATTTTAATATAGCGTTTCAAAACAGTTC 1042
D 1096 GCCGTGTTTCAAGTAAACCTGCTAATGCCATTTTAATATAGCGTTTCAAAACAGTTC 1155
QY 1043 AGTGAATTAATAGATTAAGTTAGTTTATACATTAACATCTGACATCTGAAATTCATATA 1102
D 1156 AGTGAATTAATAGATTAAGTTAGTTTATACATTAACATCTGACATCTGAAATTCATATA 1215
QY 1103 GAGATGTTCACTGCTGCTTTTCAATTTTAAGTAAATTTTGTGTTT 1146
D 1216 GAGATGTTCACTGCTGCTTTTCAATTTTAAGTAAATTTTGTGTTT 1259

RESULT 5
AAS26214
ID AAS26214 standard; cDNA; 1387 BP.
XX
AC AAS26214;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 393.
XX
KW Human; immunosuppressive; antiarthritic; ss. antirheumatic;
KW cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN MO200155322-A2.
XX
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
```

PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 23-SEP-2000; 2000US-0234977.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240950.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246509.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(PUMA-) HUMAN GENOME SCI INC.  
XX  
PA Rosen, CA, Barash SC, Ruben SM;  
XX  
PI WPI; 2001-488783/53.  
XX  
DR P-PSDB; AAU16227.  
XX  
PT New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX  
XX  
PS Claim 1; SEQ ID No 393; 980bp; English.  
XX  
CC The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunoassay assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease. Infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,





Db 2079 ACTCCACACATAGCTGTGGAGATTCAGATGATTCCTCCCTTCAAGATTGGCT 2020  
QY 575 AGTAGTAAGGGGCAAAAGCCATCTCATTTTGGTCCATTAGTCCAGCAATGAATT 634  
Db 2019 AGTAGTAAGGGGCAAAAGCCATCTCATTTTGGTCCATTAGTCCAGCAATGAATT 1960  
QY 635 TAGTGCAATTAATTTGATTTAGTAGACATATATATATGCTGAACATAATTTCTTAA 694  
Db 1959 TAGTGCAATTAATTTGATTTAGTAGACATATATATATGCTGAACATAATTTCTTAA 1900  
QY 695 TATTTCTATTCCTGTCAGACACCTTTTCTAGACAGCTCCAGATTGGAGCATGGCTCTA 754  
Db 1899 TATTTCTATTCCTGTCAGACACCTTTTCTAGACAGCTCCAGATTGGAGCATGGCTCTA 1840  
QY 755 AGAGCTTAAGCAATTTTTCATATGCTTATATATTCATATTCATATGACATCTTATA 814  
Db 1839 AGAGCTTAAGCAATTTTTCATATGCTTATATATTCATATTCATATGACATCTTATA 1780  
QY 815 ATATATTAAGCAATGATCTGTGTAACATACATCTGACAGCCAGCCATTTGCA 874  
Db 1779 ATATATTAAGCAATGATCTGTGTAACATACATCTGACAGCCAGCCATTTGCA 1720  
QY 875 AAATAAATCTTTTATATATATATCTATGATGAGCAGCAATATACACTGGGAA 934  
Db 1719 AAATAAATCTTTTATATATATATCTATGATGAGCAGCAATATACACTGGGAA 1660  
QY 935 GAAGTGAGTTTTTGTGTTATAGGTAATTTCTAGTAAACAGCTGCTGTTTTCAG 994  
Db 1659 GAAGTGAGTTTTTGTGTTATAGGTAATTTCTAGTAAACAGCTGCTGTTTTCAG 1600  
QY 995 TTACACTGTAAGCAATTTTATATATGCTTTTCAATGCTGAGGAAATAGT 1054  
Db 1599 TTACACTGTAAGCAATTTTATATATGCTTTTCAATGCTGAGGAAATAGT 1540  
QY 1055 ACAGATTAGTTAGTACATCTGACATCTGACATCTGAGATGATGAGATGTTAGT 1114  
Db 1539 ACAGATTAGTTAGTACATCTGACATCTGACATCTGAGATGATGAGATGTTAGT 1480  
QY 1115 GGTGCTTTTCATTTAAGTAATTTTGTGTT 1146  
Db 1479 GGTGCTTTTCATTTAAGTAATTTTGTGTT 1448

RESULT 7  
AAS26709/c  
ID AAS26709 standard; DNA; 8919 BP.  
XX AAS26709;  
AC  
XX  
DT 07-NOV-2001 (first entry)  
DE  
XX  
DE Human genomic DNA encoding partial novel secreted protein, Seq ID 1683.  
XX  
XX Human: immunosuppressive; antiarthritic; ds; antihemetic;  
KM cytoskeletal; cardiant; vasotropic; cerebroprotective; nootropic;  
KM neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KM vulnery; secreted protein; rheumatoid arthritis;  
KM hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KM cerebrovascular disorder; cerebral ischemia; angiogenesis;  
KM nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KM corneal infection; wound healing; epithelial cell proliferation;  
KM skin aging; food additive; preservative; antiproliferative.  
XX  
XX Homo sapiens.  
OS  
XX  
PN WO200155322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01341.  
XX  
XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0255678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI: 2001-488783/53.  
XX  
XX New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX

PS Disclosure; SEQ ID No 1683; 980bp; English.  
XX  
CC The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capability, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence is a genomic DNA encoding a partial novel secreted protein of  
CC the invention.  
Query Match 65.2%; Score 747.2; DB 22; Length 8919;  
Best Local Similarity 99.6%; Pred. No. 5,7e-166;  
Matches 749; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 395 GTTCATGAGCAAGCACTGCTCCCACTACACTTCTGTTGGACAGGCAAGCTG 454  
DB 2200 GATGAAGACAGCACTGCTCCCACTACACTTCTGTTGGACAGGCAAGCTG 2141  
QY 455 AAGTTGATGGAACAACAACATTTCTTCAACGAGACTCTGCTGCTCACTGCC 514  
DB 2140 AAGTTGATGGAACAACAACATTTCTTCAACGAGACTCTGCTGCTCACTGCC 2081  
QY 515 ACTCCCAACACTACTGCTGAGAGATTCGATGATGCTTCCCTTTCAAGATTGCT 574  
DB 2080 ACTCCCAACACTACTGCTGAGAGATTCGATGATGCTTCCCTTTCAAGATTGCT 2021  
QY 575 AGTAGTTAAGGGCAAAAGCTCATTTGCTGCTCATTAAGTCCAGAAATGAAT 634  
DB 2020 AGTAGTTAAGGGCAAAAGCTCATTTGCTGCTCATTAAGTCCAGAAATGAAT 1961  
QY 635 TATGTGATTTTGTGATTTGAGAGCACTTAATATGCTGAGAACTAATTTCTTAA 694  
DB 1960 TATGTGATTTTGTGATTTGAGAGCACTTAATATGCTGAGAACTAATTTCTTAA 1901  
QY 695 TATTTCTATTCCTGACGACCTTTCTACAGCTGCGACATTGGAGCAATGCCCTTA 754  
DB 1900 TATTTCTATTCCTGACGACCTTTCTACAGCTGCGACATTGGAGCAATGCCCTTA 1841  
QY 755 AGACCTTAAACATTTTGTGATTTGAGAGCACTTAATATGCTGAGAACTAATTTCTTAA 814  
DB 1840 AGACCTTAAACATTTTGTGATTTGAGAGCACTTAATATGCTGAGAACTAATTTCTTAA 1781  
QY 815 ATATATTAACACATGATCTGTTGATGAGACATCTGCTGAGAACCGAGCTTTGCA 874  
DB 1780 ATATATTAACACATGATCTGTTGATGAGACATCTGCTGAGAACCGAGCTTTGCA 1721  
QY 875 AATTAATCTTTTATATATTTATCTATGAGATGACGACCAATATACTCTGGGA 934  
DB 1720 AATTAATCTTTTATATATTTATCTATGAGATGACGACCAATATACTCTGGGA 1661  
QY 935 GAAGTGAGTTTGTGATTTAGTTAATTTTCTAGTAAACAGCTTCTGTTTCAG 994  
DB 1660 GAAGTGAGTTTGTGATTTAGTTAATTTTCTAGTAAACAGCTTCTGTTTCAG 1601  
QY 995 TTACACTGTAATGCAATTTAATATATGCTTTTCAATACAGTACGAAATAGT 1054

DB 1600 TTAACACGCTGATCCATTATTAATATAGGCTTTTCAACAGTTCAGTGAATAGT 1541  
1055 ACAGATTAGTGTACATACACTGCTGACATAGCATGGAATTCATAGATGTTCACT 1114  
DB 1540 ACAGATTAGTGTACATACACTGCTGACATAGCATGGAATTCATAGATGTTCACT 1481  
1115 GGTCTGCTGAGAGTGTGCTCAATATTTTGAATTTTGAATTTTGAATTTTGAATTTT 1146  
DB 1480 GGTCTGCTGAGAGTGTGCTCAATATTTTGAATTTTGAATTTTGAATTTTGAATTTT 1449

RESULT 8  
AAH07014  
ID AAH07014 standard: cDNA: 716 bp.  
AAH07014;  
26-JUN-2001 (first entry)  
Human cDNA clone (5'-primer) SEQ ID NO:3849.  
Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.  
Homo sapiens.  
EP1074617-A2.  
07-FEB-2001.  
28-JUL-2000; 2000BP-0116126.  
29-JUL-1999; 991P-0248036.  
27-AUG-1999; 990P-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX  
XX MPI: 2001-318749/34.  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 1; SEQ ID 3849; 2537PP + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to a  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primer sets are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.  
XX  
XX Sequence 716 BP; 231 A; 148 C; 171 G; 163 T; 3 other;  
SQ  
Query Match 42.9%; Score 491.2; DB 22; Length 716;  
Best Local Similarity 97.5%; Pred. No. 4, 2e-119;  
Matches 508; Conservative 0; Mismatches 11; Indels 2; Gaps 1;  
QY 83 ACCCTGGGAGCCGAGCAGCTGAGAGAGTAGAGAGCCGACACCTGACAGACACCTGCT 142  
DB 196 AGCCCTGGGAGCCGAGCAGCTGAGAGAGTAGAGAGCCGACACCTGACAGACACCTGCT 255  
QY 143 ACAAGGTCCCAATGAGCCACAGCTGCTGAGATTTTAAACTTATAGATGAGCATGATA 202  
DB 256 ACAAGGTCCCAATGAGCCACAGCTGCTGAGATTTTAAACTTATAGATGAGCATGATA 315  
QY 203 GCTGCTGAGAGAGTGTGCTCAATATTTTACATAGACACATGAGATTAAGAGAGGAGCACTA 262  
DB 316 GCTGCTGAGAGAGTGTGCTCAATATTTTACATAGACACATGAGATTAAGAGAGGAGCACTA 375  
QY 263 ACCAGCTGTATCTGAGCAAGCCACCTTAATATGGAATGGAATGCTGTTGAGGCTG 322  
DB 376 ACCAGCTGTATCTGAGCAAGCCACCTTAATATGGAATGGAATGCTGTTGAGGCTG 435  
QY 323 GATGCCCTAAATATATTTTGTGACACATGCGCTTAGTAGAGTTCAGGTCATATGTTA 382  
DB 436 GATGCCCTAAATATATTTTGTGACACATGCGCTTAGTAGAGTTCAGGTCATATGTTA 495  
QY 383 GATGCCCAACAGTTCATGAGCAAGCACTGACCTCCCAACATACATGCTGTTGAGCC 442  
DB 496 GATGCCCAACAGTTCATGAGCAAGCACTGACCTCCCAACATACATGCTGTTGAGCC 555  
QY 443 AGTGAAGTGTGAGAGTTGATGGAACAACAACATTTCTTCAACAGAACTCTGCTG 502  
DB 556 AGTGAAGTGTGAGAGTTGATGGAACAACAACATTTCTTCAACAGAACTCTGCTG 615  
QY 503 ACTGCTCAGTCCACCTCCCAACATGCTGATGGAATGGAATGCTGCTGCTGTTT 562  
DB 616 ACTGCTCAGTCCACCTCCCAACATGCTGATGGAATGGAATGCTGCTGCTGTTT 675  
QY 563 CAGAGATTGCTTACT--AGTTAAAGGGGCAAAAGTCCATTTC 601  
DB 676 CAGAGATTGCTTACT--AGTTAAAGGGGCAAAAGTCCATTTC 716

RESULT 9  
AAS75617  
ID AAS75617 standard: cDNA: 1035 bp.  
AAS75617;  
13-FEB-2002 (first entry)  
DNA encoding novel human diagnostic protein #11421.  
Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;



DR WPI: 2001-639362/73.  
 DR P-PSDB: ABG11430.  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PS Claim 1; SEQ ID NO 11421; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.

CC Sequence 1035 BP; 318 A; 219 C; 222 G; 276 T; 0 other:

Query Match 29.0%; Score 332.4; DB 23; Length 1035;  
 Best Local Similarity 98.9%; Pred. No. 2.6e-77;  
 Matches 356; Conservative 0; Mismatches 1; Indels 3; Gaps 2:

QY 15 AATGGAGGAGTGGAAATTTTGTGCGTTGGCGGTTCCCTCTCTTATAGTATGAT 74  
 DB 55 ACTGGAGGAGTGGAAATTTTGTGCGTTGGCGGTTCCCTCTCTTATAGTATGAT 114  
 QY 75 CATTCCGAGCCCTGCGGACCGACACCTGAGAGAGTGTAGTACCGGACCTGCCAGAAC 134  
 DB 115 CATTCCGAGCCCTGCGGACCGACACCTGAGAGAGTGTAGTACCGGACCTGCCAGAAC 174  
 QY 135 ACATCTGACAGAGTCCAGATGGCCAGCTCTGTGATTTTAAACCTATGTAGTACAG 194  
 DB 175 ACATCTGACAGAGTCCAGATGGCCAGCTCTGTGATTTTAAACCTATGTAGTACAG 234  
 QY 195 CATGTAGAGCTGTGAGAGAGTTTGCATATTTACTATGAGACATGATTAAGAACAC 254  
 DB 235 CATGTAGAGCTGTGAGAGAGTTTGCATATTTACTATGAGACATGATTAAGAACAC 294  
 QY 255 GGGCAGCTAACAGAGGTGTATGTGACAGAGCCACTTATATGATGAGAAATGCTGTTT 314  
 DB 295 GGGCAGCTAACAGAGGTGTATGTGACAGAGCCACTTATATGATGAGAAATGCTGTTT 354  
 QY 315 CAGGAGCTGATGCCCTTAATTA--TTTTTTGACACATGTGCTTC--TAGTGAAGTCCAGG 371  
 DB 355 CAGGAGCTGATGCCCTTAATTAATTTTTTTTGGACACATGCTTCTTAGTGAAGTCCAGG 414

RESULT 10  
 ID AAA08036 standard: cDNA; 830 BP.  
 XX AAA08036:

DT 19-JUN-2000 (first entry)

DE Human protein transport molecule (PTAM) encoding cDNA SEQ ID NO:10.

KW Human: protein transport molecule; PTAM; diagnosis; cytostatic;  
 KW antithrombotic; antidiabetic; immunosuppressant; antihypertensive;  
 KW antiallergic; antidiabetic; antipruritic; antipruritic; antipruritic;  
 KW dermatological; antineoplastic; antipruritic; antipruritic; antipruritic;  
 KW immunofluorescent; antihypertensive; protein transport regulator; cancer;  
 KW immune disorder; cell proliferative disorder; secretory disorder;  
 KW urticaria; allergy; abnormal vesicle trafficking; asthma;  
 KW autoimmune haemolytic anaemia; ss.

OS Homo sapiens.

PN WO200012703-A2.

PD 09-MAR-2000.

PF 26-AUG-1999; 99WO-US19616.

PR 27-AUG-1998; 98US-0098206.

PA (INCY) INCYTE PHARM INC.

PI Tang YF, Lai P, Bandman O, Yue H, Corley NC, Guegler KJ;  
 PI Gorgone GA, Baughn MR, Patterson C;

DR WPI: 2000-256642/22.

DR P-PSDB: AAY82318.

PT New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis  
 PS Claim 9; Page 70; 75pp; English.

CC AAA08035 to AAA08042 encode the human protein transport-associated  
 CC molecules (PTAMs) given in AAY82317 to AAY82324. The PTAMs have  
 CC cytosolic, antithrombotic, antidiabetic, immunosuppressant,  
 CC antihypertensive, antiallergic, antipruritic, antipruritic,  
 CC antihypertensive, osteopathic, dermatological, antineoplastic, antipruritic,  
 CC hepatotropic, antipruritic, antipruritic and antipruritic activities, and  
 CC regulate protein transport. PTAM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PTAM. PTAM polynucleotides are useful for  
 CC diagnosing conditions associated with PTAM, comprising detecting PTAM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the  
 CC biological sample. Diseases prevented, treated or diagnosed include cell  
 CC proliferative disorders such as cancers, immune disorders, secretory  
 CC disorders and other conditions associated with abnormal vesicle  
 CC trafficking, such as allergies, asthma, urticaria and autoimmune  
 CC haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as  
 CC a targeting or delivery mechanism for bringing pharmaceutical agents  
 CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related  
 CC disorders. PTAM, its catalytic or immunogenic fragments are useful for  
 CC drug screening using libraries of compounds. PTAM polynucleotides are  
 CC useful for generating hybridisation probes useful in mapping the  
 CC naturally occurring genomic sequences.

CC Sequence 830 BP; 198 A; 208 C; 195 G; 229 T; 0 other:

Query Match 22.0%; Score 251.6; DB 21; Length 830;  
 Best Local Similarity 72.4%; Pred. No. 4e-56;  
 Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0:

QY 162 GCTCTGATTTTAAACCTATGTATGATGATGATGATGATGATGATGATGATGATGAT 221  
 DB 114 CATCTGTGATTTTAAAGACCTATGTATGATGATGATGATGATGATGATGATGATGAT 173  
 QY 222 ATATTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 281  
 DB 174 ATGTCTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 233  
 QY 282 AGGCCACCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 341  
 DB 234 CAGCCACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 293

OY 342 TTGACATTCGCTCTAGTGTGAGTTCAGGATCATATATGTTGATGCCAACAGTTTCATG 401  
DB 294 TTGAAATGCTTCCTTCAGGAGGATCCAAATCAGCGGTGATCCAGCGCTGTTCAAG 353  
OY 402 AGCAAGCACTCAGTCCCAACTAGTCTTGTGTGACACGAGTGAAGTTTG 461  
DB 354 ATGAAAGCCACACCAAGCCAGACAGGTCCTTTGTTGATCTGATGATGATGAAAGTTTG 413  
OY 462 ATGGAACCAACACATTTCTTCAACAGAACTTCCTGACCTGACCTGACCTGACCTCCA 521  
DB 414 AGGGAAACAAACAGGAGCTTCAACAGAACTTATCTGACCGCCAGGCTTACCCA 473  
OY 522 ACATATCTGTGTGGAAGTTCAGATTCCTTCCTGTTTCAAGATTGCTTACTAGT 581  
DB 474 GCAACACAGCTGTGGAAGTTCAGATTCCTTCCTGTTTCAAGATTGCTTACTAGT 533  
OY 582 AAGGGGCAAAAGTCATTCATTCATTCGTC 611  
DB 534 GGGGTGCGAGAGGCTCTTCTTCTTCATTC 563

RESULT 11  
AAS26215  
ID AAS26215 standard; cDNA; 864 BP.  
XX AAS26215;  
AC AAS26215;  
XX AAS26215;  
DT 07-NOV-2001 (first entry)  
XX  
DE Human cDNA encoding a novel secreted protein, Seq ID 394.  
XX  
KW Human; immunosuppressive; antithrombotic; ss; antirheumatic;  
KW cytoskeletal; cardiant; vasotropic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnerability; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
XX Homo sapiens.  
OS  
XX  
PN WO200155322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01341.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0227709.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0233397.  
PR 14-SEP-2000; 2000US-0233398.  
PR 14-SEP-2000; 2000US-0233399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-0244675.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.



PI Schlegel R, Endege WO, Monahan JE;  
XX WPI: 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer  
XX  
XX Claim 1: Page 1144-1145; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (1) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (1) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 245 BP; 79 A; 47 C; 46 G; 73 T; 0 other;  
Query Match 19.2%; Score 220; DB 23; Length 245;  
Best Local Similarity 100.0%; Pred. No. 5.2e-48;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 837 GGTACTACATCTACTGCTGTAACCCAGCCCTATTGCAAAAATTAATCTTTTATATAT 896  
DB 244 GGTACTACATCTACTGCTGTAACCCAGCCCTATTGCAAAAATTAATCTTTTATATAT 185  
QY 897 TATCATGGAGTGTGACGACATATATACACTCTGGGAAGAAGTGGAGTTTGGTTAT 956  
DB 184 TATCATGGAGTGTGACGACATATATACACTCTGGGAAGAAGTGGAGTTTGGTTAT 125  
QY 957 AGGTTAATTTTCTAGTAACAAACAGTTCCTGTTTCACTTAACACATGTAATGCCATTTT 1016  
DB 124 AGGTTAATTTTCTAGTAACAAACAGTTCCTGTTTCACTTAACACATGTAATGCCATTTT 65  
QY 1017 AATATATGGCTTTTCAAAATCAGTTCACTGTAAGAAATAGTAC 1056  
DB 64 AATATATGGCTTTTCAAAATCAGTTCACTGTAAGAAATAGTAC 25  
RESULT 13  
ABV36989/C  
ID ABV36989 standard; cDNA; 441 BP.  
XX  
XX ABV36989;  
XX  
XX 16-SEP-2002 (first entry)  
XX  
XX Human prostate expression marker cDNA 36980.  
XX  
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX pharmacogenomic marker; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200160860-A2.  
XX  
XX 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
XX  
XX 17-FEB-2000; 2000US-183319P.  
XX  
XX 16-MAR-2000; 2000US-189862P.  
XX  
XX 25-MAY-2000; 2000US-207454P.

PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Endege WO, Monahan JE;  
XX WPI: 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer  
XX  
XX Claim 1: Page 7616; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (1) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (1) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 441 BP; 118 A; 95 C; 102 G; 126 T; 0 other;  
Query Match 19.2%; Score 220; DB 23; Length 441;  
Best Local Similarity 100.0%; Pred. No. 6.4e-48;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 837 GGTACTACATCTACTGCTGTAACCCAGCCCTATTGCAAAAATTAATCTTTTATATAT 896  
DB 271 GGTACTACATCTACTGCTGTAACCCAGCCCTATTGCAAAAATTAATCTTTTATATAT 212  
QY 897 TATCATGGAGTGTGACGACATATATACACTCTGGGAAGAAGTGGAGTTTGGTTAT 956  
DB 211 TATCATGGAGTGTGACGACATATATACACTCTGGGAAGAAGTGGAGTTTGGTTAT 152  
QY 957 AGGTTAATTTTCTAGTAACAAACAGTTCCTGTTTCACTTAACACATGTAATGCCATTTT 1016  
DB 151 AGGTTAATTTTCTAGTAACAAACAGTTCCTGTTTCACTTAACACATGTAATGCCATTTT 92  
QY 1017 AATATATGGCTTTTCAAAATCAGTTCACTGTAAGAAATAGTAC 1056  
DB 91 AATATATGGCTTTTCAAAATCAGTTCACTGTAAGAAATAGTAC 52  
RESULT 14  
AAS58525  
ID AAS58525 standard; cDNA; 318 BP.  
XX  
XX AAS58525;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX cDNA #1201 encoding portion of a human colon tumour protein.  
XX  
XX Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200173027-A2.  
XX  
XX 04-OCT-2001.

PF 22-MAR-2001; 2001WO-US09246.  
 XX PR 24-MAR-2000; 2000US-191597P.  
 PR 04-MAY-2000; 2000US-202024P.  
 PR 05-MAY-2000; 2000US-202189P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Meagher MJ, Xu J, King GE;  
 XX WPI; 2001-611627/70.  
 DR New colon tumour proteins and related nucleic acid, useful for  
 XX treatment, prevention, diagnosis and monitoring of cancer  
 PT Claim 4; Page 246; 299pp; English.  
 XX  
 XX Th present invention relates to the isolation of novel cDNA sequences  
 CC encoding for at least an immunogenic portion of human colon tumour  
 CC proteins. The sequences of the invention are useful in pharmaceutical  
 CC compositions and vaccines for the prevention and treatment of cancers  
 CC such as colon cancer. They are also useful for the diagnosis and  
 CC monitoring of such cancers. Antibodies to the colon tumour proteins  
 CC and antigen presenting cells that express polynucleotides encoding  
 CC colon tumour proteins can be used to inhibit the development of  
 CC cancers. T-cells that react specifically with colon tumour proteins  
 CC are useful for removing tumour cells from samples (e.g. blood) and  
 CC for cancer treatment. The polynucleotides sequences are also useful in  
 CC gene therapy. AAS57325-AAS58880 represent the cDNA sequences of the  
 CC invention that encode for portions of human colon tumour proteins.  
 XX  
 SQ Sequence 318 BP; 76 A; 82 C; 83 G; 76 T; 1 other;  
 Query Match 16.1%; Score 184.8; DB 23; Length 318;  
 Best Local Similarity 73.8%; Pred. No. 1e-38;  
 Matches 234; Conservative 0; Mismatches 83; Indels 0; Gaps 0;  
 QY 195 CATGTAGAGCTGCTGAGGAGTTTGTCAATATTTACTATGACAAATGATTAAGAAGAC 254  
 DB 1 CCTGCAGAGCTGCTGAGGAGTTTGTCAATGCTACTACACCACCATGATGAAGCGCGGC 60  
 QY 255 GGGCACTAACCGCTGATCTGGACAGGCCACCTTAATGGAATGGAATGCTGTTT 314  
 DB 61 GTTGCTGCTCCGCTGTACATGGGCAGCAGCCCTGGTCTGGAATGGCAATGCTGTTT 120  
 QY 315 CAGGCTGGATGCCCTTAATATTTTGTACACATTCCTTCTAGTGAGTTCCAGGTCA 374  
 DB 121 CAGGACAAAGAAATCCTTGAGTGAGTTTGTGAATGTTGCTTCCAGCGAGTTCCAAATCA 180  
 QY 375 ATATGTTAGATGTCGAACCACTTCATGAGCAAGCACTCACTCCCAAACTACAGTTCTTG 434  
 DB 181 CGGTGGTAGACTGCCGCTTTCATGATGAAGCCACACCAAGCCAGCACCGTCTCTTG 240  
 QY 435 TTGTGACCAAGTGAAGTTTGAATGGAACAAACAAACATTTCTTCAACAGAACT 494  
 DB 241 TTGTCATCTGGATCAGTGAGTTTGGGGGACAAACAAACAGGACCTTCAACAGAACT 300  
 QY 495 TCCTGCTGAGTGTCTAG 511  
 DB 301 TCATNCTGACCGCCAG 317  
 RESULT 15  
 ABN38056  
 ID -ABN38056 standard; DNA; 60 BP.  
 XX AC ABN38056;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX Human spliced transcript detection oligonucleotide SEQ ID NO:10804.  
 DE Human; mouse; rat; splice transcript; detection; RNA transcript;  
 KW

KW splice variant; transcriptome; oligonucleotide library; ss.  
 XX Homo sapiens.  
 OS WO200210449-A2.  
 PN 07-FEB-2002.  
 PD 20-JUL-2001; 2001WO-IB01903.  
 XX 28-JUL-2000; 2000US-221607P.  
 XX 02-MAY-2001; 2001US-287724P.  
 PR (COMP-) COMPUGEN INC.  
 PA Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
 XX WPI; 2002-257383/30.  
 DR New oligonucleotide libraries comprising oligonucleotides which  
 XX selectively hybridize to mRNAs transcribed from a transcription unit of  
 PT a genome, useful for detecting tissue-, pathology-, and  
 CC developmental-specific genes  
 CC Example 1; SEQ ID 10804; 47pp; English.  
 CC The present invention describes oligonucleotide libraries for detecting  
 CC messenger RNAs that populate a (sub-)transcriptome, where the  
 CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
 CC transcription units that populate a genome. The library comprises  
 CC several oligonucleotides, each capable of hybridising selectively to a  
 CC set of messenger RNAs transcribed from a given transcription unit of  
 CC the genome, which encodes one or more messenger RNA splice variants.  
 CC The oligonucleotide libraries are useful for detecting mRNAs from a  
 CC biological sample, in expression profiling studies, in qualitatively or  
 CC quantitatively characterising the corresponding transcriptome, and in  
 CC detecting RNA transcripts and splice variants of human or animal  
 CC transcriptomes. The libraries may also be used as specialised mini  
 CC libraries to detect transcripts of a sub-transcriptome under a  
 CC particular biological or pathological state, and so allowing the  
 CC detection of tissue- and pathology-specific genes such as those genes  
 CC only expressed in specific tissue under a specific pathological  
 CC condition; to detect developmental specific genes; and to detect RNA  
 CC transcripts and splice variants of a transcriptome of a patient suffering  
 CC from a particular disorder. ABN27253 to ABN59589 represent  
 CC oligonucleotide sequences from rats, humans and mice, which are used in  
 CC the exemplification of the present invention.  
 CC N.B. the sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 60 BP; 17 A; 17 C; 11 G; 15 T; 0 other;  
 Query Match 5.2%; Score 60; DB 24; Length 60;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-06;  
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 383 GATTGCCAACCAAGTTCATGAGCAAGCAACTCAGTCCCAAACTACAGTTCTTGTGTGACC 442  
 DB 1 GATTGCCAACCAAGTTCATGAGCAAGCAACTCAGTCCCAAACTACAGTTCTTGTGTGACC 60  
 Search completed: June 28, 2003, 01:09:33  
 Job time : 296.199 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:22:51 ; Search time 3099.3 seconds

(Without alignments)  
10761.067 Million cell updates/sec

Title: US-09-763-902B-15

Perfect score: 1146

Sequence: 1 taccctccgagagagaatggg.....tttaagtaattttgtttt 1146

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pt:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_ov:\*  
22: em\_ov:\*  
23: em\_ph:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_vl:\*  
31: em\_htg\_hum:\*  
32: em\_htg\_hum:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_pln:\*  
36: em\_htg\_pln:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_sy:\*  
40: em\_sy:\*  
41: em\_sy:\*  
42: em\_sy:\*  
43: em\_sy:\*  
44: em\_sy:\*  
45: em\_sy:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1123	98.0	2577	9	BC014888
2	1064	92.8	2692	9	AK023289
3	888	77.5	896	9	AF201942
4	747.2	65.2	178863	9	HS596C15
5	429	37.4	429	9	AF246127
6	429	37.4	429	9	HS427591
7	415	36.2	567	9	HS427591
8	301.4	26.3	148320	2	AC107438
9	301.4	26.3	173939	2	AC110357
10	301.4	26.3	186222	2	AC123243
11	274.2	23.9	232171	10	AL731672
12	251.6	22.0	939	9	BC002687
13	251.6	22.0	963	9	BC003029
14	251.6	22.0	965	9	AF156957
15	251.6	22.0	972	9	BC000759
16	251.6	22.0	1176	9	BC003410
17	251.6	22.0	1176	9	HS132813
18	250.6	21.9	1049	9	AK026360
19	245.8	21.4	186415	2	AC110699
20	243.8	21.3	644	10	AF156958
21	186.2	16.2	182118	2	AC128310
22	184.8	15.9	318	6	AX261550
23	182	15.9	135619	9	AL354809
24	182	15.9	147310	2	AL513495
25	168.4	14.7	295166	2	AL161914
26	164.4	14.3	153483	2	AC111510
27	142	12.4	207621	2	AL845495
28	120.4	10.5	175991	2	AC094606
29	76.6	6.7	203822	2	AC073698
30	73.4	6.4	173939	2	AC110357
31	61.4	5.4	186222	2	AC123243
32	55.8	4.9	203822	2	AC073698
33	55.6	4.9	507	3	AF156959
34	55.6	4.9	538	3	AV071250
35	53.2	4.6	1141	6	AX083744
36	51.8	4.5	187013	2	AC116920
37	51.2	4.5	1141	6	AX083744
38	50.6	4.4	86827	3	PFMAL3P5
39	50.6	4.4	349980	6	AX344564
40	50	4.4	310779	2	AC005140
41	49.2	4.3	17738	6	AX346441
42	48.8	4.3	1035	6	AX083745
43	48.8	4.3	66016	2	AC011162
44	48.2	4.2	29016	2	AC117266
45	48	4.2	14001	3	PF00018B

## ALIGNMENTS

RESULT 1  
BC014888  
LOCUS BC014888 2577 bp mRNA  
DEFINITION Homo sapiens, hypothetical protein P15-2, clone MGC:10178  
IMAGE:3908026, mRNA, complete cds.  
ACCESSION BC014888  
VERSION BC014888.1 GI:15928849  
KEYWORDS MGC  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 2577)  
AUTHORS Strausberg, R.  
TITLE Direct Submission

Pred. No. is the number of results predicted by chance to have a

## JOURNAL

Submitted (01-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

## REMARK

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 Info@bcsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Harris.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 9295187.

## FEATURES

Location/Qualifiers  
 1..2577  
 /organism="Homo sapiens"  
 /db\_xref="locusid:55916"  
 /db\_xref="taxon:9606"  
 /clone="MGC:10178 IMAGE:3908026"  
 /tissue\_type="uterus, leiomyosarcoma"  
 /clone\_id="NIH\_MGC\_71"  
 /lab\_host="DH10B"  
 /note="vector: PCMV-SPORT6"  
 132..560  
 /codon\_start=1  
 /product="hypothetical protein p15-2"  
 /protein\_id="AAH4888.1"  
 /db\_xref="gi:15928850"  
 /translation="MATSDFKTYVDQACHAAEFVNIYETMDKRRRLRLYLDKA  
 TLIMNGAVSGLALNNEFDLPSEVOMVDCOPHEBATSOTVVLVVGSTVKE  
 DKMKHFNPLTLAOSTPNNVMKIASCFRQDMSS"

## CDS

BASE COUNT 802 a 448 c 442 g 885 t

## ORIGIN

Query Match 98.0%; Score 1123; DB 9; Length 2577;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-244;  
 Matches 1123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

24 GTGGAATAATTTTGTGCGTTGGGGGTTTGGCTCTCTCATTAAGTATGATTCGCGCA 83  
 1 GTGGAATAATTTTGTGCGTTGGGGGTTTGGCTCTCTCATTAAGTATGATTCGCGCA 60  
 84 GCCCTGGGACCGGACGCTGAGAGTAGACGCCGACACGCGACACACACACCTGCTA 143  
 61 GCCCTGGGACCGGACGCTGAGAGTAGACGCCGACACGCGACACACACCTGCTA 120  
 144 CAAAGTCCAGATGCGCAGCTCTCTGATTTTAAACTTATGATGATCAGGATGAG 203  
 121 CAAAGTCCAGATGCGCAGCTCTGATTTTAAACTTATGATGATCAGGATGAG 180  
 204 CTGCTGAGAGTTTGTCAATATTACTATGAGACATGATGATAAGAGCGGCACATA 263  
 181 CTGCTGAGAGTTTGTCAATATTACTATGAGACATGATGATAAGAGCGGCACATA 240  
 264 CCAAGCGTATCTGAGACAGCCACTTAATATGATGATAAGAGCGGCACATA 323  
 241 CCAAGCGTATCTGAGACAGCCACTTAATATGATGATAAGAGCGGCACATA 300  
 324 ATGCCCTAAATATTTTGTGACACATTCGCTTCTAGTGAATTCAGGATCAATATGTAG 383

|||||  
 Db 301 ATGCCCTAAATATTTTGTGACACATTCGCTTCTAGTGAATTCAGGATCAATATGTAG 360  
 384 ATGCCCTAAATATTTTGTGACACATTCGCTTCTAGTGAATTCAGGATCAATATGTAG 443  
 361 ATGCCCTAAATATTTTGTGACACATTCGCTTCTAGTGAATTCAGGATCAATATGTAG 420  
 444 GTGGAATGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 503  
 421 GTGGAATGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
 504 CTGCTGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 563  
 481 CTGCTGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 564 AGATTTGCTGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 623  
 541 AGATTTGCTGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600  
 624 CAATTTGCTGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 683  
 601 CAATTTGCTGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660  
 684 AATTTCTTAAATTTTCTAATTCCTGACAGACCTTTTCTAGAGAGAGAGAGAGAGAG 743  
 661 AATTTCTTAAATTTTCTAATTCCTGACAGACCTTTTCTAGAGAGAGAGAGAGAGAG 720  
 744 ATGCCCTAAAGAGTTTAAACTATTTTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 803  
 721 ATGCCCTAAAGAGTTTAAACTATTTTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
 804 ACATTTCTTAAATTTTCTAATTCCTGACAGACCTTTTCTAGAGAGAGAGAGAGAGAG 863  
 781 ACATTTCTTAAATTTTCTAATTCCTGACAGACCTTTTCTAGAGAGAGAGAGAGAGAG 840  
 864 GCCTATTTGCAAAATAAATCTTTTAAATTTTCTAGAGAGAGAGAGAGAGAGAGAGAG 923  
 841 GCCTATTTGCAAAATAAATCTTTTAAATTTTCTAGAGAGAGAGAGAGAGAGAGAGAG 900  
 924 CACTCTGGAG 983  
 901 CACTCTGGAG 960  
 984 CCGTTTTCAGTACAGCTGTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1043  
 961 CCGTTTTCAGTACAGCTGTAATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 1044 GTGGAATAATTTTGTGCGTTGGGGGTTTGGCTCTCTCATTAAGTATGATTCGCGCA 1103  
 1021 GTGGAATAATTTTGTGCGTTGGGGGTTTGGCTCTCTCATTAAGTATGATTCGCGCA 1080  
 1104 AGATTTGCTGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146  
 1081 AGATTTGCTGAGAGTTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1123  
 RESULT 2  
 LOCUS AK023289  
 DEFINITION Homo sapiens cDNA FLJ13227 fls, clone OVARC100071, weakly similar to Homo sapiens NTF2-related export protein NTF1 (NTF1) mRNA.  
 ACCESSION AK023289  
 VERSION AK023289.1 GI:10435160  
 KEYWORDS oligo capping; fls (full insert sequence).  
 SOURCE Homo sapiens ovary, tumor tissue cDNA to mRNA, clone\_id:OVARC1  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 Isegaki, T., Ota, T., Hayashi, K., Sugiyama, T., Osuki, T., Suzuki, Y., Nishikawa, T., Nagai, K., Sugano, S., Shitatori, A., Sudo, H., Magatsuna, M., Hosoiri, T., Kaku, Y., Kodaira, H., Kondo, H.,





```

/codon_start=1
/product="DC9"
/protein_id="AA086878.1"
/db_xref="GI:9295188"
/translation="MATSLEPKTYVDQACRAEEFVNIYETMDKRRRLRLYLKDA
TLINGNVAVGLALNFDLTPSESEQVNMDCQPVHEQATOSQTVLVTSQYKF
DGNKHFNFNLITLQSTPNNTWKIASCFRQDSSS"
BASE COUNT      267 a      177 c      178 g      274 t
ORIGIN

```

```

Query Match      77.5% Score 888; DB 9; Length 896;
Best Local Similarity 100.0%; Pred. No. 9.2e-191;
Matches 888; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 11 AGAAGATGGAGGGGTGAAAAATTTTGTGCGTTGGCGGGTTCCCTCTTCATAAGTAT 70
DB 1 AGAAGATGGAGGGGTGAAAAATTTTGTGCGTTGGCGGGTTCCCTCTTCATAAGTAT 60
OY 71 TGATCATTCGCGACCCCTGGCGGACCGGACAGCTGAGGAGTGTGAGCGGACGCTGCA 130
DB 61 TGATCATTCGCGACCCCTGGCGGACCGGACAGCTGAGGAGTGTGAGCGGACGCTGCA 120
OY 131 GAACACACTGCTACAGAGTCCAGATGCGACGCTCTGTGATTTTAAACTTATGTAGAT 190
DB 121 GAACACACTGCTACAGAGTCCAGATGCGACGCTCTGTGATTTTAAACTTATGTAGAT 180
OY 191 CAGCATGTAGAGTGTGAGGAGTGTGCAATTTTACTATGAGACATGTGATTAAGA 250
DB 181 CAGCATGTAGAGTGTGAGGAGTGTGCAATTTTACTATGAGACATGTGATTAAGA 240
OY 251 AGACGGGACACTAACGAGCTGTATCTGGACCAAGCCACTTAATATGATGAAATGCT 310
DB 241 AGACGGGACACTAACGAGCTGTATCTGGACCAAGCCACTTAATATGATGAAATGCT 300
OY 311 GTTTCAGGCGTGGATGCCCTTAATATTTTGTGACACATTCCTTGTAGTGTCCAG 370
DB 301 GTTTCAGGCGTGGATGCCCTTAATATTTTGTGACACATTCCTTGTAGTGTCCAG 360
OY 371 GTTCATATGTATAGTATGCAACAGCTTATATGACAGCACTAGTCCCAACTACAGT 430
DB 361 GTTCATATGTATAGTATGCAACAGCTTATATGACAGCACTAGTCCCAACTACAGT 420
OY 431 CTGTGTGACAGAGTGAAGTGTGATGGAAGCAACAACATTTTCTTAACACAG 490
DB 421 CTGTGTGACAGAGTGAAGTGTGATGGAAGCAACAACATTTTCTTAACACAG 480
OY 491 AACTTCTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
DB 481 AACTTCTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
OY 551 TGTCTCCGTTTCAAGATGTGTGATGATTAAGGGGCAAAAGTCCATTTCTATTGCT 610
DB 541 TGTCTCCGTTTCAAGATGTGTGATGATTAAGGGGCAAAAGTCCATTTCTATTGCT 600
OY 611 CCATTAGTTCAGCAATGGAATTTATGTAATTTTGTGATGTAAGAGCACTATATA 670
DB 601 CCATTAGTTCAGCAATGGAATTTATGTAATTTTGTGATGTAAGAGCACTATATA 660
OY 671 TGTGCTGAAGTAAATTTCTTAATATTTTCTGATGCTGATGACACCTTTCTAGCAGCT 730
DB 661 TGTGCTGAAGTAAATTTCTTAATATTTTCTGATGCTGATGACACCTTTCTAGCAGCT 720
OY 731 GCCAGTTTGAGAGTATGCTCTTAAGAGCTTAAACATTTTATTAATGCTTATATA 790
DB 721 GCCAGTTTGAGAGTATGCTCTTAAGAGCTTAAACATTTTATTAATGCTTATATA 780
OY 791 CATTCCTACTAATGACATCTTATATATATTAACACATGATCTGTGATACATACT 850
DB 781 CATTCCTACTAATGACATCTTATATATATTAACACATGATCTGTGATACATACT 840
OY 851 CACTGTGAGACCCAGCTATGCAAAAAATTAATCTTTTATTAATATTA 898
DB 841 CACTGTGAGACCCAGCTATGCAAAAAATTAATCTTTTATTAATATTA 888

```

```

RESULT 4
HS596C15/
LOCUS
DEFINITION
Human DNA sequence from clone 596C15 on chromosome Xq23. Contains
the GUCY2F gene for guanylate cyclase 2F, retinal (RC 4,6,1,2,
RENGC-2, Rod Outer Segment Membrane Guanylate Cyclase 2F, ROS-GC2,
GC-F) and a potentially alternatively spliced novel gene. Contains
ESTs, an STS and GSSs, complete sequence.
ACCESSION
AL031387
VERSION
HG: GC-F; GUCY2F; RENGC-2; retinal guanylate cyclase 2F; Rod Outer
Segment Membrane Guanylate Cyclase; ROS-GC2.
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1. (bases 1 to 178863)
Howden, P.
Direct Submission
Submitted (01-APR-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On Mar 24, 1999 this sequence version replaced gi:4464241.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence is the entire insert of clone 596C15. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
596C15 is from the library RPI4 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR: pcyrac2.
FEATURES
source
1. 178863
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q23"
/clone="RP4-596C15"
/clone_11b="RPI4-4"
519..560
/note="21 copies 2 mer ac 95% conserved"
1079..1353
/note="match: GSS A0132738"
1626..1893
/note="AluB3 repeat: matches 1..283 of consensus"
1963..3662
/note="L1P45 repeat: matches 4450..6143 of consensus"
3683..4070
/note="L1P repeat: matches 2885..3261 of consensus"
complement(4444..12256)
/gene="djs96C15.1"
complement(join(4444..6630,7518..7662,11006..11092,
11867..>11941))
/gene="djs96C15.1"
/product="djs96C15.1.1 (novel protein) (isoform 1)"
/note="match: ESTs AA233523 AA354023 AA333479 AA400543
N27567 AA831754 AA459145 AA328722 R67870 H00878 AA148904
AA26578 N20297 AA782964 AA400435 AA458927 AA526010
A1268152 AA234442 AA148905 A1066541 A1467944 R67871 R25673

```



Db 6517 ACTCCCAACATACCTCTGTGGAGATTGCAAGTGATGCTCCGTTTCAAGATTGGTCT 6458  
QY 575 AGTAGTTAAAGGGGCAAAAGTCCATCTCTCATTTGGTCCATTAGTTCACAGCAATGAATTT 634  
Db 6457 AGTAGTTAAAGGGGCAAAAGTCCATCTCTCATTTGGTCCATTAGTTCACAGCAATGAATTT 6398  
QY 635 TAGTGAATTTATTTGATGTAGAGCAGCACTAATATATGCTGGAACAACTAAATTTCTTTAA 694  
Db 6397 TAGTGAATTTATTTGATGTAGAGCAGCACTAATATATGCTGGAACAACTAAATTTCTTTAA 6338  
QY 695 TATTTTCTATCTCTGCAGCACCTTTCTAGCAGCTGCCAGTTTGAGCATTGCCCTCTA 754  
Db 6337 TATTTTCTATCTCTGCAGCACCTTTCTAGCAGCTGCCAGTTTGAGCATTGCCCTCTA 6278  
QY 755 AGAGCTTTAAACTATTTTACATGCTTATATACATTTCCACTAATGACATTTCTTTATA 814  
Db 6277 AGAGCTTTAAACTATTTTACATGCTTATATACATTTCCACTAATGACATTTCTTTATA 6218  
QY 815 ATAATATTAACACATGATCTTGGTACTAACATCTACTGTGAACCCAGCCTATTGCAA 874  
Db 6217 ATAATATTAACACATGATCTTGGTACTAACATCTACTGTGAACCCAGCCTATTGCAA 6158  
QY 875 AAATAAAATCTTTTATATATATCTATGATGTCAGCACAATATACACTCTGGGAA 934  
Db 6157 AAATAAAATCTTTTATATATATCTATGATGTCAGCACAATATACACTCTGGGAA 6098  
QY 935 GAAGTGGAGTTTTTGGTTATTAGGTTAATTTCTAGTAAACACAGTTGCCCTGTTTTCAG 994  
Db 6097 GAAGTGGAGTTTTTGGTTATTAGGTTAATTTCTAGTAAACACAGTTGCCCTGTTTTCAG 6038  
QY 995 TTAACACTGGTAAATGCCATTTTAAATATATGCGTTTTTCAAACTCAGTTCAGTGAATACT 1054  
Db 6037 TTAACACTGGTAAATGCCATTTTAAATATATGCGTTTTTCAAACTCAGTTCAGTGAATACT 5978  
QY 1055 ACAGATTTAGGTTTACATACTACTGACATCTGCAATTCGATATGATAGATGTTTCACT 1114  
Db 5977 ACAGATTTAGGTTTACATACTACTGACATCTGCAATTCGATATGATAGATGTTTCACT 5918  
QY 1115 GGTGCTTTTTCATTTTAAAGTAAATTTTGTGTTT 1146  
Db 5917 GGTGCTTTTTCATTTTAAAGTAAATTTTGTGTTT 5886

RESULT 5  
AF246127  
LOCUS Homo sapiens p15-2 mRNA, complete cds. PRI 21-JUN-2000  
DEFINITION AF246127  
ACCESSION AF246127.1 GI:8575519  
VERSION AF246127.1  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Kang, Y. and Cullen, B.  
TITLE p15-2, a homologous protein of p15, interacts with Tap  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 429)  
AUTHORS Kang, Y. and Cullen, B.  
TITLE Direct Submission  
JOURNAL Submitted (15-MAR-2000) Genetics, Duke University, 427 CARL  
Building, DMC 3025, Durham, NC 27705, USA  
FEATURES  
source  
1. 429  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1. 429  
/codon\_start=1  
/product="p15-2"  
/protein\_id="AAF78034.1"  
/db\_xref="GI:8575520"  
/translation="MATSLDFKTYVDQACRAAEFVNYYETMDKRRRLRLYLDKA

TLIINGNAVSGLDALNNFFDTLPSEFOVNMDCOPVHEQATQSQTTVLVVTSGTVKE  
DGNKQHFQNFLLTAQSTPNNTVKNKIASDCERFQDWSS"  
BASE COUNT 125 a 85 c 94 g 125 t  
ORIGIN  
Query Match 37.4%; Score 429; DB 9; Length 429;  
Best Local Similarity 100.0%; Pred. No. 8.9e-87;  
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 155 ATGGCCACGCTCTCTGGATTTTAAACTTATGTAGATCAGGCATGTAGAGCTCTCTGAGGAG 214  
Db 1 ATGGCCACGCTCTCTGGATTTTAAACTTATGTAGATCAGGCATGTAGAGCTCTCTGAGGAG 60  
QY 215 TTGTCAATATTTACTATGAGCAATGGATAAAAGACGGGCACCTAACCAAGGCTGTAT 274  
Db 61 TTGTCAATATTTACTATGAGCAATGGATAAAAGACGGGCACCTAACCAAGGCTGTAT 120  
QY 275 CTGGACAGGCCACCTTATATGGAATGGAATGCTGTTTCAGGGCTGGATGCCCTAAAT 334  
Db 121 CTGGACAGGCCACCTTATATGGAATGGAATGCTGTTTCAGGGCTGGATGCCCTAAAT 180  
QY 335 AATTTTGTGACACATTTGCTTCTAGTGAGTTCCAGTGCAATATGTAGATGCCCAACCA 394  
Db 181 AATTTTGTGACACATTTGCTTCTAGTGAGTTCCAGTGCAATATGTAGATGCCCAACCA 240  
QY 395 GTTCATGACAGCAACTCAGTCCCAAACTACAGTTCTTTGTGTGACCAAGTGAAGTGTG 454  
Db 241 GTTCATGACAGCAACTCAGTCCCAAACTACAGTTCTTTGTGTGACCAAGTGAAGTGTG 300  
QY 455 AGTTTGTGAGGAAACAAACAACTTCTTCAACCAAGTTCCTGCTGACTGCTCAGTCC 514  
Db 301 AGTTTGTGAGGAAACAAACAACTTCTTCAACCAAGTTCCTGCTGACTGCTCAGTCC 360  
QY 515 ACTCCCAACAACTACTGTTGGAGATTGCAAGTATTGCTTCCGTTTCAAGATTGGTCT 574  
Db 361 ACTCCCAACAACTACTGTTGGAGATTGCAAGTATTGCTTCCGTTTCAAGATTGGTCT 420  
QY 575 AGTAGTTAA 583  
Db 421 AGTAGTTAA 429  
RESULT 6  
HSA277591  
LOCUS Homo sapiens mRNA for p15-2a protein (p15-2 gene). PRI 19-NOV-2000  
DEFINITION HSA277591  
ACCESSION AJ277591  
VERSION AJ277591.1 GI:8920231  
KEYWORDS p15-2 gene; p15-2a protein.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Herold, A., Suyama, M., Rodrigues, J.P., Braun, I.C., Kutay, U.,  
Carmona-Fonseca, C., Bork, P. and Izaurralde, E.  
TITLE TAP (NXF1) belongs to a multigene family of putative RNA export  
factors with a conserved modular architecture  
JOURNAL Mol. Cell. Biol. 23, 8996-9008 (2000)  
REFERENCE 2 (bases 1 to 429)  
AUTHORS Izaurralde, E.  
TITLE Direct Submission  
JOURNAL Submitted (26-APR-2000) Izaurralde E., Gene Expression, EMBL,  
Meyerhofstrasse 1, 69117 Heidelberg, GERMANY  
COMMENT Related sequence: AJ278323  
FEATURES  
Location/Qualifiers  
1. 429  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xq23"  
/tissue\_type="testis"  
1. 429  
gene

CDS

/gene="p15-2"  
1..429  
/function="p15-2"  
/note="the p15-2 gene produces two alternatively spliced  
mRNAs encoding for p15-2a and p15-2b proteins"  
/evidence="experimental  
/product="p15-2a protein"  
/protein\_id="CA86337.1"  
/db\_xref="GI:8920232"  
/translation="MATSIDFKTYVDACRAAEFVNIYETMDKRRRLRLYLKRA  
TLINGNVAVSGLDALNNEFDLPSEFQVNMDCOPHEQATQSQTIVLVTSVKR  
DGNKQHFNFNELLTAOSTPNNVTWMTIASDCFRFDWSSS"  
BASE COUNT 125 a 85 c 94 g 125 t  
ORIGIN

Query Match 37.4%; Score 429; DB 9; Length 429;  
Best Local Similarity 100.0%; Pred. No. 8.9e-87;  
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 ATGGCCACGCTCTCGATTTTAAACCTTATGATCAGCGATGATGAGCTGCTGAGAG 214  
DB 1 ATGGCCACGCTCTCGATTTTAAACCTTATGATCAGCGATGATGAGCTGCTGAGAG 60  
QY 215 TTGTCAATATTTACTATGACACATGATGATAAAGAGAGCGGACCTAACCGAGCTGTAT 274  
DB 61 TTGTCAATATTTACTATGACACATGATGATAAAGAGAGCGGACCTAACCGAGCTGTAT 120  
QY 275 CTGACACAGGCGACCTTAATATGAAATGGAATGCTTTTCAGGCTGATGCCCTAAT 334  
DB 121 CTGACACAGGCGACCTTAATATGAAATGGAATGCTTTTCAGGCTGATGCCCTAAT 180  
QY 335 AATTTTGTGACATATGCTCTTCTAGAGAGTCCAGCTCATATGTTTATGATGCCACCA 394  
DB 181 AATTTTGTGACATATGCTCTTCTAGAGAGTCCAGCTCATATGTTTATGATGCCACCA 240  
QY 395 GTTCATGACGACGACCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 454  
DB 241 GTTCATGACGACGACCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 300  
QY 455 AAGTTTATGGAACCAACCAACATTTCTCAACGAGAACTTCGCTGCTGCTGCTGCTC 514  
DB 301 AAGTTTATGGAACCAACCAACATTTCTCAACGAGAACTTCGCTGCTGCTGCTGCTC 360  
QY 515 ACTGCCACACATATGCTGTGTGAGATGCAAGTATGCTTCCGTTTCAAGATTGCTCT 574  
DB 361 ACTGCCACACATATGCTGTGTGAGATGCAAGTATGCTTCCGTTTCAAGATTGCTCT 420  
QY 575 AGTAGTTAA 583  
DB 421 AGTAGTTAA 429

RESULT 7  
LOCUS HSA278323 567 bp mRNA linear PRI 19-NOV-2000  
DEFINITION Homo sapiens mRNA for p15-2b protein (p15-2 gene).  
ACCESSION AJ278323  
VERSION AJ278323.1 GI:9663146  
KEYWORDS p15-2 gene; p15-2b protein.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 567)  
Harold, A., Suyama, M., Rodrigues, J.P., Braun, I.C., Kutay, U.,  
Carro-Fonseca, C., Bork, P. and Izaurralde, E.  
TAP (NHEI) belongs to a multigene family of putative RNA export  
factors with a conserved modular architecture  
Mol. Cell. Biol. 23, 8996-9008 (2003)  
JOURNAL 2 (bases 1 to 567)  
REFERENCE 1  
AUTHORS Izaurralde, E.

TITLE Direct Submission  
JOURNAL Submitted (06-JUN-2000) Izaurralde E., Gene Expression, EMBL,  
Meyerhofstrasse 1, 69117 Heidelberg, GERMANY  
COMMENT Related sequence: AJ277591.  
FEATURES  
source  
1..567  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xq23"  
/tissue="testis"  
1..567  
/gene="p15-2"  
1..567  
/function="p15-2"  
/note="the p15-2 gene produces two alternatively spliced  
mRNAs encoding for p15-2a and p15-2b proteins"  
/evidence="experimental  
/product="p15-2b protein"  
/protein\_id="CAC01129.1"  
/db\_xref="GI:9663147"  
/translation="VTNHPALCTPAGRPRFARAPGTPSLVSSRFLIPRTLSWEI  
ARGDVGDFKTYVDACRAAEFVNIYETMDKRRRLRLYLKRA TLINGNVAVSGL  
DALNNEFDLPSEFQVNMDCOPHEQATQSQTIVLVTSVKR DGNKQHFNFNELL  
TAOSTPNNVTWMTIASDCFRFDWSSS"  
BASE COUNT 146 a 131 c 142 g 148 t  
ORIGIN

Query Match 36.2%; Score 415; DB 9; Length 567;  
Best Local Similarity 100.0%; Pred. No. 1.3e-83;  
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 GGATTTTAAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 228  
DB 153 GGATTTTAAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 212  
QY 229 CTATGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 288  
DB 213 CTATGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 272  
QY 289 CTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 348  
DB 273 CTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 332  
QY 349 ATGCGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408  
DB 333 ATGCGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 392  
QY 409 AACTCAGTCCCAAACTACAGTTCTTGTGTGACGAGTGAAGTGAAGTGAAGTGAAGTGA 468  
DB 393 AACTCAGTCCCAAACTACAGTTCTTGTGTGACGAGTGAAGTGAAGTGAAGTGAAGTGA 452  
QY 469 CAACACATATTTCTTCAACGACGACTCTCTGACGCTGACGCTGACGCTGACGCTGAC 528  
DB 453 CAACACATATTTCTTCAACGACGACTCTCTGACGCTGACGCTGACGCTGACGCTGAC 512  
QY 529 TGTGTGAGAGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 583  
DB 513 TGTGTGAGAGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 567

RESULT 8  
LOCUS AC107438 148320 bp DNA linear HTG 13-JUN-2002  
DEFINITION Rattus norvegicus clone CH230-160A14, \*\*\* SEQUENCING IN PROGRESS  
ACCESSION AC107438  
VERSION AC107438.3 GI:21736312  
KEYWORDS HTG; HTGS; PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

1 (bases 1 to 148320)

Alzury, D. M., Adams, C., Adio-Ofodola, B., Ali-Osman, F. R., Banks, T. C., Allbrook, S. L., Amarantunga, H. C., Are, J. R., Ayele, M., Bait, K., Barberia, J., Benton, J., Blmage, K., Blankenburg, K., Bonini, D., Bouck, J., Bowle, S., Britve, M., Brown, E., Brown, M., Bryant, N. P., Buah, C., Burch, P., Burkett, C., Burrill, K. L., Byrd, N. C., Caron, T. F., Carter, M., Cavazos, S. R., Chacko, J., Chaves, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C. D., Cox, C., Coyle, M., Dathorne, S. R., David, R., Davila, M. L., Davis, C., Day-Carroll, L., Dederich, D. A., Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H., Douthwaite, R. J., Draper, H., Dugan-Rocha, S., Dudin, K. T., Earnhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Escotto, M., Falls, T., Ferrazuto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gortelli, J. H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hultzy, S., Hume, J., Jackson, L. E., Jacobson, B., Jais, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Kovvah, J., Kovar, C., Kratochvic, J., Kuneshe, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichtensterg, O., Lieu, C., Liu, J., Liu, W., Loulegeed, H., Lozard, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mathwney, E., McLeod, M. P., Meador, M., Mel, G., Metzker, M., Miner, G., Miner, Z., Mitchell, J., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokekwo, S., Oguh, M., Okunolu, G., Oragunye, N., Oyedero, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pul, L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojoubokan, I., Rolle, M., Ruiz, S., Severy, G., Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalton, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

## Direct Submission

2 (bases 1 to 148320)

Worley, K.C.

Direct Subr

Submitted

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 148320)

Worley, K.C.

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

On Jul 12,

Genome Center

Center: Baylor College of Medicine

Center code: BCM

web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hqsc-help@bcm.tmc.edu](mailto:hqsc-help@bcm.tmc.edu)

----- Project Information -----

Center project name: GMZC

Center clone name: CH230-160A14

## Summary Statistics

Chemistry: Dye-terminator Bld Dye: 100% of reads

Assembly program: Phrab: version 0.990329

Assembly program: final, version 0:000000  
Consensus quality: 89845 bases at least 040

Consensus quality: 94598 bases at least 030

Consensus quality: 98393 bases at least 20 consensus quality: 94550 bases at least 20

```
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hpsc.bcm.tmc.edu/docs/genbank\_drift.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

1	1578:	contig of 1578	bp in length
1579	1678:	gap of unknown length	
1679	3151:	contig of 1473	bp in length
3152	3251:	gap of unknown length	
3252	4752:	contig of 1501	bp in length
4753	4852:	gap of unknown length	
4853	5878:	contig of 1026	bp in length
5879	5978:	gap of unknown length	
5979	7250:	contig of 1272	bp in length
7251	7350:	gap of unknown length	
7351	8428:	contig of 1078	bp in length
8429	8528:	gap of unknown length	
8529	10032:	contig of 1504	bp in length
10033	10132:	gap of unknown length	
10133	11243:	contig of 1111	bp in length
11243	11343:	gap of unknown length	
11344	12929:	contig of 1586	bp in length
12930	13029:	gap of unknown length	
13030	14302:	contig of 1273	bp in length
14303	14402:	gap of unknown length	
14403	15468:	contig of 1066	bp in length
15468	15568:	gap of unknown length	
15569	17010:	contig of 1442	bp in length
17011	17110:	gap of unknown length	
17111	18317:	contig of 1207	bp in length
18318	18417:	gap of unknown length	
18418	19702:	contig of 1285	bp in length
19703	19802:	gap of unknown length	
19803	21600:	contig of 1558	bp in length
21361	21460:	gap of unknown length	
21461	23143:	contig of 1683	bp in length
23144	23343:	gap of unknown length	
23244	24991:	contig of 1748	bp in length
24992	25091:	gap of unknown length	
25092	26564:	contig of 1473	bp in length
26565	26664:	gap of unknown length	
26665	27783:	contig of 1119	bp in length
27784	27883:	gap of unknown length	
27884	29648:	contig of 1765	bp in length
29649	29748:	gap of unknown length	
29749	31751:	contig of 2003	bp in length
31752	31851:	gap of unknown length	
31852	33098:	contig of 1247	bp in length
33099	33198:	gap of unknown length	
33199	34820:	contig of 1622	bp in length
34821	34920:	gap of unknown length	
34921	36154:	contig of 1234	bp in length
36155	36244:	gap of unknown length	
36255	37964:	contig of 1710	bp in length
37965	38064:	gap of unknown length	
38065	39211:	contig of 1147	bp in length
39212	39311:	gap of unknown length	
39312	41243:	contig of 1932	bp in length
41244	41343:	gap of unknown length	
41344	42978:	contig of 1635	bp in length
42979	43078:	gap of unknown length	
43079	44146:	contig of 1068	bp in length
44147	44246:	gap of unknown length	
44247	45967:	contig of 1721	bp in length
45968	46067:	gap of unknown length	
46068	47786:	contig of 1729	bp in length
47787	47886:	gap of unknown length	



Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peterson, L., Pichens, R., Primus, E., Pu, L.L., Quiles, M., Ren, X.,  
Rivas, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Scherer, S., Scott, G., Shen, H., Shoochari, N., Sisson, I.,  
Soderren, E., Sotnik, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Swatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, D., Zorrilla, S., Nelson, D.,  
Wenstock, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 173939)  
Worley, K.C.

Direct Submission  
Submitted (11-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 173939)  
Worley, K.C.

Direct Submission  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced g1:18847131.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GRM  
Center clone name: CH230-276m2  
----- Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 136387 bases at least Q40  
Consensus quality: 139690 bases at least Q30  
Consensus quality: 142385 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 49 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1030: contig of 1030 bp in length  
1031 1130: gap of unknown length  
1131 2654: contig of 1524 bp in length  
2655 2754: gap of unknown length  
2755 3833: contig of 1079 bp in length  
3834 3933: gap of unknown length  
3933 5317: contig of 1383 bp in length  
5317 5416: gap of unknown length  
5416 6880: contig of 1464 bp in length  
6881 7987: gap of unknown length  
7987 8087: contig of 1007 bp in length  
8088 9455: gap of unknown length  
9455 9553: gap of 1368 bp in length  
9553 10645: contig of 1090 bp in length  
10646 11918: contig of 1173 bp in length  
11919 12018: gap of unknown length  
12019 13599: contig of 1581 bp in length  
13600 13699: gap of unknown length

13700 15310: contig of 1611 bp in length  
15311 15410: gap of unknown length  
15411 17265: contig of 1855 bp in length  
17266 17366: gap of unknown length  
17366 19287: contig of 1821 bp in length  
19287 19287: gap of unknown length  
19287 21646: contig of 2359 bp in length  
21646 23514: gap of unknown length  
23514 23515: contig of 1769 bp in length  
23515 26113: gap of unknown length  
26113 26214: contig of 2499 bp in length  
26214 26214: gap of unknown length  
26214 28949: gap of 2736 bp in length  
28949 29050: gap of unknown length  
29050 30792: contig of 1743 bp in length  
30793 30892: gap of unknown length  
30893 33813: contig of 2821 bp in length  
33814 33913: gap of unknown length  
33914 37076: gap of 3163 bp in length  
37076 37176: contig of 3163 bp in length  
37176 39759: gap of unknown length  
39759 39860: gap of 2583 bp in length  
39860 42880: gap of unknown length  
42880 42980: gap of 3020 bp in length  
42980 45805: gap of unknown length  
45805 45905: gap of 2825 bp in length  
45905 45904: gap of unknown length  
45904 4601: contig of 3697 bp in length  
4601 49701: gap of unknown length  
49701 51721: contig of 2020 bp in length  
51721 51822: gap of unknown length  
51822 54941: gap of 3120 bp in length  
54941 55042: gap of unknown length  
55042 57798: contig of 2757 bp in length  
57798 57899: gap of unknown length  
57899 61347: contig of 3449 bp in length  
61347 61448: gap of unknown length  
61448 65731: gap of 4284 bp in length  
65731 65832: gap of unknown length  
65832 67836: contig of 2005 bp in length  
67836 67936: gap of unknown length  
67936 70267: gap of 2330 bp in length  
70267 70366: gap of unknown length  
70366 74979: gap of 4613 bp in length  
74979 75079: gap of unknown length  
75079 78340: contig of 3261 bp in length  
78340 78440: gap of unknown length  
78440 82207: contig of 3767 bp in length  
82207 82307: gap of unknown length  
82307 85016: contig of 2709 bp in length  
85016 85116: gap of unknown length  
85116 88120: contig of 3004 bp in length  
88120 88220: gap of unknown length  
88220 91247: gap of 3027 bp in length  
91247 91347: gap of unknown length  
91347 94541: contig of 3194 bp in length  
94541 100271: gap of unknown length  
100271 100371: gap of 5630 bp in length  
100371 105497: gap of unknown length  
105497 105597: gap of 5126 bp in length  
105597 111059: gap of unknown length  
111059 11159: gap of unknown length  
11159 115817: contig of 4658 bp in length  
115817 115917: gap of unknown length  
115917 121770: gap of 5853 bp in length  
121770 121870: gap of unknown length  
121870 127414: contig of 5544 bp in length  
127414 127514: gap of unknown length  
127514 135104: contig of 7590 bp in length  
135104 145776: gap of unknown length  
145776 145876: contig of 10572 bp in length  
145876 145877: gap of unknown length  
145877 145877: contig of 8991 bp in length



\* 154868 154967: gap of unknown length  
 \* 154968 162848: contig of 7881 bp in length  
 \* 162849 162948: gap of unknown length  
 \* 162949 173939: contig of 10991 bp in length.  
 Location/Qualifiers  
 source Rattus norvegicus  
 1..173939  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 /clone="CH230-27682"

BASE COUNT 50612 a 33701 c 33123 g 48110 t 8393 others  
 ORIGIN

Query Match 26.3% Score 301.4; DB 2; Length 173939;  
 Best Local Similarity 73.2%; Pred. No. 3e-58;  
 Matches 532; Conservative 0; Mismatches 161; Indels 34; Gaps 10;

395 GTTCATGAGCAAGCAAGCAAGCTCCCAACATCAAGTCTTGTGTGACCAAGTGAAGTGTG 454  
 66941 GCTAAGAGCAAGCTACCCAGTACAGTACAGTGTCTGTGTGACCAAGTGAAGTGTG 67000  
 455 AAGTTGATGCAAGCAAGCAAGCTTCTCAACCAAGTCTGTGTGACCTGTGCTGCTG 514  
 67001 AAGTTGATGCAAGCAAGCAAGCTTCTCAACCAAGTCTGTGTGACCTGTGCTGCTG 67060  
 515 ACTCCCAACCAAGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 574  
 67061 ACTCCCAACCAAGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 67120  
 575 AGTATTTAAGGGGCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 634  
 67121 AGTATTTAAGGGGCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 67178  
 635 TA-----TGATATTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 680  
 67179 TAAATGATATTTAAGTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 67238  
 681 CTAATTTCTTATTA-TTTCTATCTCTGTGACAGCTTTCTG--CAGTGTGCAATT 737  
 67239 CTAATTTCTTATTA-TTTCTATCTCTGTGACAGCTTTCTG--CAGTGTGCAATT 67298  
 738 TGAAGCATTTGCTCTGAGAGCTTTAAACATTTTAACTATTTTAACTATTTTAACTATTT 795  
 67299 TGAAGCATTTGCTCTGAGAGCTTTAAACATTTTAACTATTTTAACTATTTTAACTATTT 67357  
 796 CACATTAAGCATTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 855  
 67358 CACATTAAGCATTTCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 67415  
 856 TGAACCCAGCTATTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 915  
 67416 TGAACCCAGCTATTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 67474  
 916 CAATATTAACACTCTGAGAGAGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 975  
 67475 CAATATTAACACTCTGAGAGAGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 67534  
 976 ACAGCTGCTCTGTTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1035  
 67535 ACAGCTGCTCTGTTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 67595  
 1036 TCACTTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1095  
 67586 TATAGATTTATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 67645  
 1096 GCATATA 1102  
 67646 GCATATA 67652

RESULT 10  
 AC123243 186222 bp DNA linear HTG 17-JUL-2002  
 LOCUS AC123243  
 DEFINITION Rattus norvegicus clone CH230-31p19, \*\*\* SEQUENCING IN PROGRESS

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS

\*\*\*, 65 unordered pieces.  
 AC123243  
 AC123243.2 GI:21745987  
 HTG: HTGS PHASE1.  
 Norway rat.  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 186222)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
 Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayale, M., Banks, T.,  
 Barbarta, J., Benton, J., Bimaye, K., Blankenburg, K., Bonin, D.,  
 Bouck, J., Bowler, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Bunney, C., Burch, P., Burkett, C., Burdell, K.L., Byrd, N.C.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Dumbin, K.J.,  
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Correll, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hognes, M., Hollway, C., Hollins, B.,  
 Homani, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,  
 Karlsson, E., Kelly, S., Khan, U., Kling, L., Korvah, J., Kovat, C.,  
 Kratochovic, J., Kuresh, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martinale, A., Martinez, E.,  
 Massey, E., McWhiney, E., McLeod, M.P., Meador, M., Mel, G., Metzger, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabdet, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okunodu, N.,  
 Orazungue, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,  
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Swalek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleciyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Glibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 186222)  
 Direct Submission  
 Worley, K.C.  
 Direct Submission  
 Submitted (29-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 186222)  
 Direct Submission  
 Worley, K.C.  
 Direct Submission  
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 14, 2002 this sequence version replaced gi:21240072.  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: G1T  
 Center clone name: CH230-31p19



## Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 123845 bases at least Q40

Consensus quality: 128005 bases at least Q30

Consensus quality: 131716 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html))

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 65 contigs. The true order of the pieces

\* is not known and their order in this sequence record 1

\* arbitrary. Gaps between the contigs are represented as

\* runs of  $N$ , but the exact sizes of the gaps are unknown

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
\* be provided

\* be preserved  
\*

1	1633	1632	contig of 1632 bp in length
*	1733	1732	gap of unknown length
*	3450	3449	contig of 1717 bp in length
*	3550	4717	contig of 1168 bp in length
*	4718	4817	gap of unknown length
*	5982	5981	contig of 1164 bp in length
*	6082	7502	gap of unknown length
*	7503	7602	contig of 1421 bp in length
*	7603	9031	gap of unknown length
*	9132	9131	contig of 1429 bp in length
*	10391	10390	contig of 1259 bp in length
*	10391	10490	gap of unknown length
*	10491	11746	contig of 1256 bp in length
*	11747	11846	gap of unknown length
*	11847	13117	contig of 1871 bp in length
*	13718	13817	gap of unknown length
*	13818	15195	contig of 1378 bp in length
*	15196	15295	gap of unknown length
*	15296	16597	contig of 1102 bp in length
*	15398	16697	gap of unknown length
*	16498	17899	contig of 1402 bp in length
*	17900	17999	gap of unknown length
*	18000	19371	contig of 1372 bp in length
*	19372	19471	gap of unknown length
*	19472	20716	contig of 1245 bp in length
*	20717	20816	gap of unknown length
*	20817	22597	contig of 1781 bp in length
*	22598	22697	gap of unknown length
*	22698	24044	contig of 1347 bp in length
*	24045	24144	gap of unknown length
*	24145	25711	contig of 1327 bp in length
*	25472	25571	gap of unknown length
*	25572	26916	contig of 1345 bp in length
*	26917	27016	gap of unknown length
*	27017	28825	contig of 1809 bp in length
*	28826	28925	gap of unknown length
*	28926	30260	contig of 1335 bp in length
*	30361	30460	gap of unknown length
*	30461	31852	contig of 1392 bp in length
*	31853	33152	gap of unknown length
*	33153	33709	contig of 1757 bp in length
*	33710	33809	gap of unknown length
*	33810	35330	contig of 1521 bp in length
*	35331	35330	gap of unknown length
*	35431	36771	contig of 1541 bp in length
*	36972	37071	gap of unknown length
*	37072	39696	contig of 2625 bp in length
*	39697	39796	gap of unknown length
*	39797	41922	contig of 2126 bp in length
*	41923	42022	gap of unknown length
*	42023	43620	contig of 1598 bp in length
*	43621	44530	gap of unknown length
*	44530	45330	contig of 1610 bp in length

45531	45430: gap of unknown length
45531	47622: contig of 2192 bp in length
47523	47722: gap of unknown length
47723	49854: contig of 2132 bp in length
49855	49954: gap of unknown length
49955	51438: contig of 1484 bp in length
51439	51538: gap of unknown length
51539	51780: contig of 2242 bp in length
53811	53880: gap of unknown length
53881	55569: contig of 1689 bp in length
55570	55669: gap of unknown length
55570	58232: contig of 2563 bp in length
58333	58332: gap of unknown length
58333	60649: contig of 2317 bp in length
60650	60749: gap of unknown length
60750	63599: contig of 2850 bp in length
63600	63699: gap of unknown length
63700	65434: contig of 1735 bp in length
65343	65534: gap of unknown length
65535	66538: contig of 3004 bp in length
66839	66638: gap of unknown length
68639	70893: contig of 2255 bp in length
70894	70993: gap of unknown length
70994	73637: contig of 2644 bp in length
73638	73737: gap of unknown length
75980	75980: contig of 2243 bp in length
75981	76080: gap of unknown length
76081	79351: contig of 3271 bp in length
79352	79451: gap of unknown length
79452	81755: contig of 2304 bp in length
81756	81855: gap of unknown length
81856	84787: contig of 2932 bp in length
84788	84887: gap of unknown length
84888	87647: contig of 2760 bp in length
87648	87747: gap of unknown length
87748	91057: contig of 3310 bp in length
91058	91157: gap of unknown length
91158	94696: contig of 3539 bp in length
94697	94796: gap of unknown length
94797	96545: contig of 1749 bp in length
96546	96645: gap of unknown length
96646	100902: contig of 4257 bp in length
100903	101002: gap of unknown length
101003	105287: contig of 4285 bp in length
105288	105387: gap of unknown length
105388	108272: contig of 2885 bp in length
108273	108372: gap of unknown length
108373	112332: contig of 3860 bp in length
112333	112332: gap of unknown length
112333	114916: contig of 2584 bp in length
114917	115016: gap of unknown length
115017	116825: contig of 3809 bp in length

Query Match	26.38;	Score 301.4;	DB 2;	Length 186222;
-------------	--------	--------------	-------	----------------

Best Local Similarity 73.28; Pred. No. 3e-58;

Matches 532; Conservative 0; Mismatches 161; Indels 34; Gaps 10;

OY	395	GTTCTAGGCAAGCAACACTCAGTCCCAAACCTAACAGTCTTGTTGTGCACAGTGAACCTGTG	454
Db	148228	GCTAAGAAGCGAAGCTACCACAGATACAGTACAGTCTCGTGTGACCGAGTGAGTGCTG	148165
OY	455	AAGTTTGSATGGAAACAACAACATTCTTCACACGAACTTCCTCGTGAATGCTCATGCTC	514
Db	148168	AAGTTTGTGGAAACAACAACACTCTTTTAACACGAACTTCCTGTTGACTGCGCACAGC	148109
OY	515	ACTGCCAACAATPACTGCTGTGGAGATPAGCACTGATGATGCTTCGCTTTCOAAGATGCTC	574
Db	148108	ACTCTTAACACACGCGTGTGGAGATPAGCACTGATGATGCTTCGCTTTCOAAGATGCTC	148049
OY	575	AGTAGTAAAGGGGCAAAAAGTCCATTTCATTTGGTCATTAGTTCACACAATTTGAATT	634
Db	148048	AGTATTTAAAGAGCAAAAGCCCCG--TCGATTGTCATTAGTTCACACAAGAGATT	147991

635 TA-----TGGAATTAATTTGATGTAGAGCACTATA-----TATGCTGAAA 680  
 147990 TAAAGTGAATTAATGATTTGTTCCAAAGCACTATAAAGTATGATGCTAAA 147931  
 681 CTAATATTTCTTAATA-TTTCTATCTGTCAGCAGCTTTTCAG--CAGCTGCCAGT 737  
 147930 CTAATATGATTAATTAATTTTATTTTCTAGACAGCAGCTTTGCTAGATTAAGTGCAGCT 147871  
 738 TGGAGCAATGGCCCTAGAGCTTTAAACATATTTTACATGCTTAT--ATACATTC 795  
 147870 TGGATTAATGCCC-TAAACCTTTAAAGATGATTTTAAATGCTTATGCTATGCAATTC 147812  
 796 CACATATGACATTTCTTAATATATTAATTAACATGATCTGCTACTAATACACAGC 855  
 147811 CACATGATGATCTGCTCAAGTAATTTTA--ACATGATCTAGTACTAGATACACAGC 147754  
 856 TGAACCCAGCCTATTTGCAAAATTAATCTTTTATTAATTAATTAATGATGATGACCA 915  
 147753 TGAACCCAGCCTATTTGCAAAATGATGATGATGATGATGATGATGATGATGATGAT 147695  
 916 CAAATATACACTCTGGGAG 975  
 147694 CAAATATACACTCTGGGAG 147635  
 976 ACAGCTGCTCTGTTTTCAGTTAAGCACTGATGATGATGATGATGATGATGATGATGATGAT 1035  
 147634 AGCAATATGCTCTGTTTTCAGTTAAGCACTGATGATGATGATGATGATGATGATGATGATGAT 147584  
 1036 TCAGCTCAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1095  
 147583 TATAGATTTTGAAGTCAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 147524  
 1096 GCATATA 1102  
 147523 GCATATA 147517

RESULT 11  
 AL731672 232171 bp DNA 1linear ROD 20-AUG-2002  
 LOCUS Mouse DNA sequence from clone RP23-161L11 on chromosome X, complete  
 DEFINITION sequence.  
 ACCESSION AL731672  
 VERSION AL731672.11 GI:22415912  
 KEYWORDS HMG.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 232171)  
 Clark, S.  
 Direct Submission  
 Submitted (20-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Aug 21, 2002 this sequence version replaced g1:2195555.  
 COMMENT Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one MJ3 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-161L11 is from the RP23-Mouse PAC Library constructed by the group of Pletzer de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACe3.6.

FEATURES  
 source  
 1. 232171  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /clone="RP23-161L11"  
 /clone\_1lb="RP23-161L11"

BASE COUNT 71072 a 46436 c 45947 g 68716 t  
 ORIGIN

Query Match 23.9%; Score 274.2; DB 10; Length 232171;  
 Best Local Similarity 72.6%; Pred. No. 4,1e-52;  
 Matches 526; Conservative 0; Mismatches 143; Indels 56; Gaps 11;

395 GTTCATGAGCAAGCAATCAGTCCCAACTACAGTCTGTTGTGACAGTGGAGCTG 454  
 45590 GCTAAGAGCAAGCAATCAGTCCCAACTACAGTCTGTTGTGACAGTGGAGCTG 45649  
 455 AAGTTGATGAGCAAGCAATCAGTCTGTTGTGACAGTCTGTTGTGACAGTGGAGCTG 514  
 45650 AAGTTGATGAGCAAGCAATCAGTCTGTTGTGACAGTCTGTTGTGACAGTGGAGCTG 45709  
 515 ACTCCCAATATCTGTTGTGAGAGATGCAAGTATGCTTCCGTTTCAAGATTGGCT 574  
 45710 ACTCCCAATATCTGTTGTGAGAGATGCAAGTATGCTTCCGTTTCAAGATTGGCT 45769  
 575 AGTATTAAGGAGGCAAGTATGCTTCCGTTTCAAGATTGGCT 634  
 45770 AGTATTAAGGAGGCAAGTATGCTTCCGTTTCAAGATTGGCT 45827  
 635 TATGTGAATTAATTTGAT-----GTGAAGCACTATTAATTTGCTGAAGC 681  
 45828 AGAGAGATTAATTAATTTGCTGAAGCACTATTAATTTGCTGAAGC 45887  
 682 TAAATTTCTTAATTAATTTGCTGAAGCACTATTAATTTGCTGAAGC 741  
 45888 TAAATTTCTTAATTAATTTGCTGAAGCACTATTAATTTGCTGAAGC 45947  
 742 GCATGCGCTCTGAAGCTTAAACCTA--TTTTTTTACATGCTTAT--ATACATTCGA 797  
 45948 TTAATGCGCTCTGAAGCTTAAACCTA--TTTTTTTACATGCTTAT--ATACATTCGA 46006  
 798 CTAATGACATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 857  
 46007 CTACGACACCTCTCAAGTAATTTTA--ACATGATCTAGTACTAGTATTAATTAATTAATTAAT 46064  
 858 AACCCAGCCTATTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 917  
 46065 AACCCAGCCTATTTGCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 46121  
 918 ATATACACTCTGGGAG 977  
 46122 AGATGAGCTCTG-----AGATGAGCTCTGAGTATTAATTAATTAATTAATTAATTAATTAAT 46172  
 978 AGGTGCGCTGTTTCAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1037  
 46173 AC-ATGCGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 46225  
 1038 AGTTGAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1097  
 46226 TGTTC-----ACTAGTTTAAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 46270

QY 1098 ATATA 1102  
 Db 46271 ATRTA 46275  
 RESULT 12  
 BC002687  
 LOCUS  
 DEFINITION BC002687 939 bp mRNA linear PRI 12-JUL-2001  
 Homo sapiens, NTF2-related export protein 1, clone MGC:3469  
 IMAGE:3608037, mRNA, complete cds.  
 ACCESSION BC002687  
 VERSION BC002687.1 GI:12803702  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 939)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CNLA Library Preparation: Rubin Laboratory  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: [http://www.nisc.nih.gov/nisc\\_mgc/nhgrl.nih.gov](http://www.nisc.nih.gov/nisc_mgc/nhgrl.nih.gov)  
 Contact: Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M.,  
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
 Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Snyder, B., Stantipop, S., Thomas, P.J.,  
 Tlonsong, E.E., Touchman, J.W., Tsurganov, C., Vogt, J.L., Walker, M.A.,  
 Zhang, L.-H. and Green, E.D.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 12 Row: m Column: 11  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 7019470.  
 Location/Qualifiers  
 1. 939  
 /organism="Homo sapiens"  
 /db\_xref="LocustID:29107"  
 /db\_xref="taxon:9606"  
 /clone="MGC:3469 IMAGE:3608037"  
 /tissue\_type="uterus, endometrium adenocarcinoma"  
 /clone\_lib="NIH-MGC-44"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOT87"  
 252. 674  
 /codon\_start=1  
 /product="NTF2-related export protein 1"  
 /protein\_id="AAH02687.1"  
 /db\_xref="GI:12803703"  
 /translation="MASVDFKTYVDACRAEEFVNYVYTTMDKRRRLSLRYMGAT  
 LVNAGNAVSGESISEFEMLPSSEFOISVDCOPVDEPTPQTVLVVIGCSYVFE  
 GNRQDFNMFITTAQSPENITWIKISDCFRQDMAS"  
 BASE COUNT 217 a 253 c 250 g 219 t  
 ORIGIN  
 Query Match 22.0%; Score 251.6; DB: 9; Length 939;  
 Best Local Similarity 72.4%; Pred. No. 1.2e-46;  
 Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 162 CGTCTGGATTATTAACCTTAGTATCAGCATGTAGAGCTGTGAGAGTTGTCA 221  
 Db 256 CATCTGTGATTTCAAGACCTATGTGATGAGCTGTGACAGCTGTGAGAGTTGTCA 315  
 QY 222 ATATTACTATGAGCAATGATTAAGAGAGAGCGGACCTAACAGGCTGTATCTGACA 281  
 Db 316 ATGTCTACTACACACACCATGATAGAGCGGGGCTTGTGCTCCCGCTGTATGAGCA 375  
 QY 282 AGGCCACCTTAATATGATGATGAAATGCTGTTTCAGGCGCTGAGTCCCTTAATATTTT 341  
 Db 376 CAGCCACCCCTGGTCTGGAATGGAATGCTGTTTCAGAGCAAGATCTTGAGAGAGTTT 435  
 QY 342 TTGACACATGCTGCTTATGAGTGTGACGATCAATATGATGATGATGATGATGATGAT 401  
 Db 436 TTGAATATGTTGCTTCCAGGAGATTTCCAAATCAGCTGTGATGATGATGATGATGAT 495  
 QY 402 AGCAACCACTAGTCCCAACATCAATGCTGTGTGTGACCACTGATGATGATGATGAT 461  
 Db 496 ATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 555  
 QY 462 ATGGAACCAACACATTTTTCACACGATCTTCTGCTGATGCTCATCTCATCTCCA 521  
 Db 556 AGGGGAACAACAACAGGAGCTTACACCAATCTCATCTGACCGCCAGCCCTACACCA 615  
 QY 522 ACAATCTGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 581  
 Db 616 GCAACACAGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 675  
 QY 582 AAAGGGCAAAAGTCCATCTCATTTGTC 611  
 Db 676 GGGGGTGGCAGAGTCTCTTGTCTATTC 705  
 RESULT 13  
 BC003029  
 LOCUS  
 DEFINITION BC003029 963 bp mRNA linear PRI 12-JUL-2001  
 Homo sapiens, NTF2-related export protein 1, clone MGC:4329  
 IMAGE:2820775, mRNA, complete cds.  
 ACCESSION BC003029  
 VERSION BC003029.1 GI:12804338  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 963)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-FEB-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: DCTD/DTF  
 CNLA Library Preparation: Rubin Laboratory  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 Info@bcsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Leticia Hsiao, Martin Kirzysinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,  
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Marra.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 9 Row: p Column: 1

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7019470.

## FEATURES

source  
1.963  
/organism="Homo sapiens"  
/db\_xref="locusid:29107"  
/db\_xref="taxon:9606"  
/clone="MGC:4329 IMAGE:2820775"  
/tissue="lung, small cell carcinoma"  
/clone\_1lb="NH\_MGC\_7"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
276..698  
CDS  
/codon\_start=1  
/product="NRF2-related export protein 1"  
/protein\_id="AAH03029.1"  
/db\_xref="gi:12804339"  
/translation="MASVDEKTYVDQACRAAEFVNYVTMDKRRLLSLRYGTAT  
LVNAGNVSQGESLSEPFEMLPSEFQISVDDQPVDEATPSQTTVLVYICGSVKRE  
GNKRDENQNFILTAQSPSNTYWKIASDCFRPDNAS"

BASE COUNT 220 a 263 c 256 g 224 t  
ORIGIN

Query Match 22.0%; Score 251.6; DB 9; Length 963;  
Best Local Similarity 72.4%; Pred. No. 1.2e-46;  
Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

162 CGCTCTGATTTTAAACTATAGTATAGCATGAGCATGATGAGCTGAGAGATTGTCA 221  
|||||  
280 CATCTGATGATTTCAAGACCTATGATGATGAGCTGAGAGATTGTGTCA 339  
|||  
222 ATATTCTATGACATGATGATTAAGAAGCGGCACTAACAGGCTGATCTGACA 281  
|||||  
340 ATGCTACTACACCATCAGATGATGAGGCGGCTTTGCTGCCGCTGATCATGAGCA 399  
|||||  
282 AGGCACCTTAATATGATGATGATGATGAGGCTGATGATGATGATGATGAT 341  
|||||  
400 CAGCACCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 459  
|||||  
342 TTGACACATTTGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 401  
|||||  
460 TTGAAATGTTGCTTCCAGGAGTTCCAAATCAGCGTGTGAGTGCAGCGCTGTATG 519  
|||||  
402 AGCAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 461  
|||||  
520 ATGAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 579  
|||||  
462 ATGAAACAAACAAACATTTCTTCAACGAGATCTGCTGCTGCTGCTGCTGCTGCT 521  
|||||  
580 AGGGAAACAAACAAACGAGCTTCAACGAGATCTGCTGCTGCTGCTGCTGCTGCT 639  
|||||  
522 ACAATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 581  
|||||  
640 GCAACACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 699  
|||||  
582 AAAGGGCAAAAGTCCATTTCTCATTTGCTC 611  
|||||  
700 GGGGGTGGCAGAGGCTCTTTGCTTCATTC 729

RESULT 14  
AF156957  
LOCUS 965 bp mRNA linear PRI 30-NOV-1999  
DEFINITION Homo sapiens NRF2-related export protein NXL1 (NXL1) mRNA, complete cds.  
ACCESSION AF156957  
VERSION AF156957.1 GI:5880864  
KEYWORDS  
ORGANISM Homo sapiens.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 965)

## AUTHORS

Black, B.E., Levesque, L., Holaska, J.M., Wood, T.C. and Paschal, B.M.  
Identification of an NRF2-related factor that binds Ran-GTP and  
regulates nuclear protein export

## REFERENCE

2 (bases 1 to 965)  
Black, B.E. and Paschal, B.M.

## TITLE

Direct Submission  
Submitted (09-JUN-1999) Center for Cell Signaling, University of  
Virginia, 7161 Hospital West, Box 577 Health Sciences Center,  
Charlottesville, VA 22908, USA

## FEATURES

source  
1.965  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1.965  
/gene="NXL1"  
272..694  
/function="stimulates nuclear protein export"  
/note="similar to the sequence presented in GenBank  
Accession Number AA85905; nuclear transport factor; binds  
Ran-GTP"  
/codon\_start=1  
/product="NRF2-related export protein NXL1"  
/protein\_id="AAD54942.1"  
/db\_xref="gi:5880865"  
/translation="MASVDEKTYVDQACRAAEFVNYVTMDKRRLLSLRYGTAT  
LVNAGNVSQGESLSEPFEMLPSEFQISVDDQPVDEATPSQTTVLVYICGSVKRE  
GNKRDENQNFILTAQSPSNTYWKIASDCFRPDNAS"

BASE COUNT 207 a 265 c 263 g 230 t  
ORIGIN

Query Match 22.0%; Score 251.6; DB 9; Length 965;  
Best Local Similarity 72.4%; Pred. No. 1.2e-46;  
Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

162 CGCTCTGATTTTAAACTATAGTATAGCATGAGCATGATGAGCTGAGAGATTGTCA 221  
|||||  
276 CATCTGATGATTTCAAGACCTATGATGATGATGATGATGATGATGATGATGAT 335  
|||||  
222 ATATTCTATGACATGATGATTAAGAAGCGGCACTAACAGGCTGATCTGACA 281  
|||||  
336 ATGCTACTACACCATCAGATGATGATGATGATGATGATGATGATGATGATGAT 395  
|||||  
282 AGGCACCTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 341  
|||||  
396 CAGCACCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 455  
|||||  
342 TTGACACATTTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 401  
|||||  
456 TTGAAATGTTGCTTCCAGGAGTTCCAAATCAGCGTGTGAGTGCAGCGCTGTATG 515  
|||||  
402 AGCAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 461  
|||||  
516 ATGAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 575  
|||||  
462 ATGAAACAAACAAACATTTCTTCAACGAGATCTGCTGCTGCTGCTGCTGCTGCT 521  
|||||  
576 AGGGAAACAAACAAACGAGCTTCAACGAGATCTGCTGCTGCTGCTGCTGCTGCT 635  
|||||  
522 ACAATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 581  
|||||  
636 GCAACACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 695  
|||||  
582 AAAGGGCAAAAGTCCATTTCTCATTTGCTC 611  
|||||  
696 GGGGGTGGCAGAGGCTCTTTGCTTCATTC 725

RESULT 15  
BC000759

LOCUS	BC000075	972 bp	mRNA	linear	PRI 12-JUL-2001
DEFINITION	Homo sapiens, NTF2-related export protein 1, clone MGC:2701				
ACCESSION	IMAGE:2820775, mRNA, complete cds.				
VERSION	BC000075				
KEYWORDS	BC000075.1 GI:12653930				
SOURCE	MGC.				
ORGANISM	Homo sapiens.				
	Homo sapiens				

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:25:38 ; Search time 78.8745 Seconds

(without alignments)  
5987.487 Million cell updates/sec

Title: us-09-763-902b-15  
Perfect score: 2013

Sequence: 1 taccctccggagagagatggg.....ttttaagtaattttgtttt 1146

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+.n2p.model -DEV=x1p  
-Q=/cgn2.1/USPTO/spool/US09763902/r/unat.24062003.130341.12453/app.query.fasta.1.2254  
-DB=SPREMBL.21 -OFMT=fasta -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCT=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIN=200 -THR\_SCORE=PCT -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09763902.qcgn.1.1.232.qrunat.24062003.130341.12453 -NCP=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database:

SPREMBL.21:  
1: sp\_archaea:  
2: sp\_bacteria:  
3: sp\_fungi:  
4: sp\_human:  
5: sp\_invertebrate:  
6: sp\_mammal:  
7: sp\_mhc:  
8: sp\_organelle:  
9: sp\_phage:  
10: sp\_plant:  
11: sp\_rodent:  
12: sp\_virus:  
13: sp\_vertebrate:  
14: sp\_unclassified:  
15: sp\_virus:  
16: sp\_bacteriophage:  
17: sp\_archaeop:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	167	8.3	200	Q9HDY6
				Q9hdY6 schizosacch

Result No.	Score	Query Match length	ID	Description
2	163.5	8.1	167	Q8T2G7
3	150.7	7.5	127	Q8T2D5
4	127.5	6.3	125	Q96VN3
5	121.5	6.0	123	Q9P8H0
6	120	6.0	146	Q8R032
7	114	5.7	688	Q917J5
8	114	5.7	690	Q9NH72
9	114	5.7	690	Q9VFR4
10	107.5	5.3	522	Q9LPR0
11	105	5.2	122	Q9E2K4
12	102.5	5.1	130	Q9VJ85
13	96	4.8	946	Q9SEV5
14	95	4.7	1378	Q95YF0
15	94	4.7	537	Q21351
16	94	4.7	646	Q8X0B8
17	94	4.7	1052	Q49676
18	90.5	4.5	428	Q9LWX6
19	90.5	4.5	2151	Q9QD48
20	87.5	4.3	460	Q9FME2
21	87	4.3	1225	Q8T237
22	86	4.3	555	Q8VNR1
23	86	4.3	1595	Q9N9S1
24	84	4.2	822	Q54923
25	83.5	4.1	427	Q9LDP9
26	83.5	4.1	596	Q9NTA6
27	83.5	4.1	757	Q8T8G9
28	83.5	4.1	865	Q9N9S1
29	83.5	4.1	2151	Q9N9T5
30	83	4.1	304	Q13628
31	83	4.1	644	Q9AC71
32	83	4.1	872	Q9J5B4
33	83	4.1	872	Q9J5B4
34	82.5	4.1	1253	Q59704
35	82.5	4.1	1454	Q66928
36	82.5	4.1	2151	Q9E127
37	82	4.1	280	Q8XNA7
38	82	4.1	507	Q9TK07
39	82	4.1	749	Q13433
40	82	4.1	753	Q8V5A9
41	81.5	4.0	579	Q9E1Y3
42	81.5	4.0	613	Q9MA69
43	81.5	4.0	815	Q9QRO3
44	81.5	4.0	6620	Q96AN2
45	81	4.0	334	Q9M255

## ALIGNMENTS

### RESULT 1

Q9HDY6 PRELIMINARY; PRT; 200 AA.  
ID Q9HDY6  
AC Q9HDY6  
DT 01-MAR-2001 (TREMBL) 16, Created)  
DT 01-MAR-2001 (TREMBL) 16, Last sequence update)  
DT 01-JUN-2002 (TREMBL) 21, Last annotation update)  
DE Putative nuclear export factor.  
GN SPAB1A10.03.  
OS Schizosaccharomyces pombe (Fission Yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes;  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972H-;  
RA O'Neil S., Harris D., Wood V., Rajandream M.A., Barrell B.G.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL512562; CAC21476.1;  
DR InterPro: IPR002075; NTF2.  
DR Pfam: PF02136; NTF2; 1.  
SQ SEQUENCE 200 AA; 23029 MW; B05E47D62B27B7FF CRC64;

### Alignment Scores:

Pred. No.: 2.09e-09 Length: 200  
Score: 167.00 Matches: 33  
Percent Similarity: 53.39% Conservative: 30  
Best Local Similarity: 27.97% Mismatches: 47  
Query Match: 8.30% Indels: 8  
DB: 3 Gaps: 3  
US-09-763-902b-15 (1-1146) x Q9HDY6 (1-200)  
OY 185 GTAGATAGGAGATAGAGAGTGTGATGATTTTACATATTACTATGAGACAATGGAT 244  
DB 1 MetcunserValLysTyrAlaGlnGluPheValGlnArgTyrSerSerLeuAsp 20  
OY 245 AAAAGAGACGGGACGACACAGGCTGATCTGACAGACAGGACCATTAATGAAATGA 304  
DB 21 ThrsmrgrasnGlyLeuAlaGluPheTyrArgLysInsSerLeuIleLeuTyrasnGly 40  
OY 305 AATGCTGTTCAGGGCTGGATGCCCTAAATATTTTGTGACACATTCCTTGTAGAG 364  
DB 41 LysProMetGln---ValThrGluPheThrSerMetIleValAsnLeuProTyrSerLys 59  
OY 365 TTCACAGTCATATGATGATGTCACACCACTGATGAGACAGACTCCCAACT 424  
DB 60 ThrLysValGluAspPheAspSerGln-----GlnValMetGlyAsnAspMet 75  
OY 425 ACAGTCTCTTGTGACACAGTGAAGTGTGATGAGAACAAACAACATTTCTTC-484  
DB 76 AsnIleIleIleValIleValSerGlyThrIleArgPheAspGlyLysLysProIleValPhe 95  
OY 485 AACCAAGACCTTCGCTGACCTGCTCAGTCCACCTCCCAACATCTGTGTGAG 538  
DB 96 SerTyrArg-----SerAlaThrThrGlnProAspPheThrIlePheLys 110  
RESULT 2  
OBT2G7  
ID 08T2G7 PRELIMINARY; PRT: 167 AA.  
AC 08T2G7  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DR 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical 19.0 kDa protein.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,  
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium".  
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AC115607; AAL92371.1;  
KW Hypothetical protein.  
SQ SEQUENCE 167 AA; 19039 MW; E16EA5FAD400BAA CRC64;  
Alignment Scores:  
Pred. No.: 4.92e-09 Length: 167  
Score: 163.50 Matches: 40  
Percent Similarity: 44.96% Conservative: 18  
Best Local Similarity: 31.01% Mismatches: 68  
Query Match: 8.12% Indels: 3  
DB: 5 Gaps: 3  
US-09-763-902b-15 (1-1146) x Q8T2G7 (1-167)  
OY 176 AAAACTATGATCATGAGCATGTAGAGCTGTGAGAGATTGTC---AATATTACTAT 232  
DB 38 LysLysIleValGlnThrIleThrProArgAlaGluValPheIleLeuGluPheTyr 57  
OY 233 GAGACATGATGATAAAGAGACGGGACCTGATACCAAGCTGTATCTGGACAAGGCCACTTA 292  
DB 58 ProLysTyrAspSerSerArgAlaAspLeuIleGlyLeuTyrLysAspHisSerValSer 77

OY 293 ATATGATGGAATGCTGTTCAGGGCTGAGTCCCTAAATATTTTGTGACACATG 352  
DB 78 IleTyrasnGlyThrGluCysLysGlyProGluHisIleGlyLysLeuValGluIle 97  
OY 353 CCTTCTGTAGTACGTCACGATCAATATGTAGATGCCACACACATGATGAGCAACT 412  
DB 98 ProAsnSerValHisValAlaGluThrPheAspAlaGlnProValProSerAsp---Asp 116  
OY 413 CAGGCCCAACTACAGTCTGTGTGACGACGAGTGAAGTGTGATGAGAAACAA 472  
DB 117 LysGluAsnProAsnIleLeuIleThrAlaThrGlyValIleTyrLysThrSer 136  
OY 473 CAACATTTCTCAACAGAACATCTCTGCTGACCTGACCTGACCTCCACCAACTACTGTG 532  
DB 137 GlnHisGlnPheHisGlnThrPheLeuValLysAspProThr---AsnSerAsnLeu 155  
OY 533 TGAAGATGCAAGTATGCTCCGT 559  
DB 156 PheTyrLeuSerTyrAspCysIleArg 164  
RESULT 3  
OBT2D5  
ID 08T2D5 PRELIMINARY; PRT: 127 AA.  
AC 08T2D5  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DR 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Putative nuclear transport factor similar to nuclear transport factor 2 (NTR2) from.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AX4;  
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,  
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium".  
RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AC115613; AAM10776.1;  
SQ SEQUENCE 127 AA; 14317 MW; 2D45B1B20021B298 CRC64;  
Alignment Scores:  
Pred. No.: 1.43e-07 Length: 127  
Score: 150.00 Matches: 37  
Percent Similarity: 46.83% Conservative: 22  
Best Local Similarity: 29.37% Mismatches: 55  
Query Match: 7.45% Indels: 12  
DB: 5 Gaps: 4  
US-09-763-902b-15 (1-1146) x Q8T2D5 (1-127)  
OY 191 CAGCATGTAGAGCGTGGAGGATGTGTCATATTACTATGAGACAATGATAAAGA 250  
DB 7 GlnValValGlyValGlyLysGlnPheValGlnHisTyrTyrGlyIlePheAspSerAsn 26  
OY 251 AGACGGGACCTACACAGGCTGATCTGACAGACGACCTTAATATGAAATGATGCT 310  
DB 27 ArgAlaGlyLeuThrGlnIleTyrGlnGlnGlnThrIleThrIleThrIleTyrGlyPhe 46  
OY 311 GTTTCAGGGCTGGATGCCCTAAATATTTTGTGACATTTGCCCT---TCTAGTAG 364  
DB 47 LeuSerGlyAlaAspAlaIleValIleHisIleValGluLeuProPheGlnGlnThrAsn 66  
OY 365 TTCACAGTCATATGATGATGTCACACCACTGATGAGACAACTCAGTCCCAACT 424  
DB 67 ArgLysIleAsnSerIleAspCysGlnGlnThrTyr-----GlnPro 80  
OY 425 ACAGTCTTGTGTGACACAGTGAAGTGAAGTGTGATGAGAAACAACAACATTTCTTC 484  
DB 81 GlyIleMetIleThrValThrGlyThrLeuIleIleAspGlyGluAlaLys----- 97

```

OY 465 AACCG---AACTCTCTGCTACTGCTCAGTCACACCCCAACAATCTGTGGAACTT 541
    |||||  |||:::  |||||  ::::  ::::
Db 98 ASngInLeuLs/PpheValGlnValPheAsnLeuAlaSerAsnGlySerPheLeuL 117
OY 542 GCAAGCTGATTGCTTCGGT 559
    ::|||  |||||
Db 118 IleAsnAspPheArg 123

RESULT 4
O96VYN3 PRELIMINARY; PRT; 125 AA.
AC O96VYN3
Dt 01-DEC-2001 (TREMBLrel. 19, Created)
Dt 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
Dt 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Nuclear transport factor 2.
GN NTF2.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; Emericella.
CX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA Espeso E.A., Penalba M.A.;
RT "Nuclear transport factor 2 from Aspergillus nidulans.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY038983; AAK71467.1;
DR InterPro: IPR002075; NTF2.
DR Pfam: PF02136; NTF2".1
SQ SEQUENCE 125 AA; 13871 MW; 8FC18BDPFC597E2F2 CRC64;

```

Alignment Scores:		4.29e-05	Length:	125
Pred. No.:	Score:	127.50	Matches:	36
Percent Similarity:	47.11%	Conservative:	21	
Best Local Similarity:	29.75%	Mismatches:	53	
Query Match:	6.33%	Indels:	11	
DB:	3	Gaps:	4	

US-09-763-902B-15 (1-1146) x Q96VN3 (1-125)

QY	206	GCTGAGAGATTGTCATATATTACTATGAGACAATGGATAAAGAAGCGGCACCTAAC	265
		:	
Db	8	AlaGlnGlnPheValThrPheTyrTyrGlnThrPheAspGlyAsnArgAlaGlyLeuAla	27
QY	266	AGCGTGTCTTCGACAAAGCCACCTTAATATGTAATGGAATGCAATGCTTTTCAGGGCTGGAT	325
Db	28	ProLeuTyrArgAspHisSerMetLeuThrPheGlnThrSerValIleGlnValAla	47
QY	326	GCCCTAAATTAATTTTGTGACATTTGCT-----TTAGTGAAGTCCAGGCAATATG	379
Db	48	GlyIleIleGlnLysLeuThrSerLeuProPheGlnLysValGlnHisGlnValSerThr	67
QY	380	TTAGATTGCCAACCAAGCTTCATGAGCAAGCACTAGCTCCCAACATCAAGTCTTGTTGGT	439
Db	68	LeuAspAlaGlnProSerGlyGlnHisGly-----GlyIleLeuValLeu	82
QY	440	ACCAATGGAACGTGAAGTTTATGTGAACAACAACATTTCTTCACACGAAGATCTCGT	499
Db	83	ValThrGlyAlaLeuLeuValAspGlnGluLysAsn-----ProMetAsnTyrThr	99
QY	500	CTGACTGCTCAGTCCACTCCCAACATACT---GTGTGAGAGATTGACAGTATGCTTC	556
Db	100	GlnThrPheGlnLeuMetProAspGlyAlaGlySerTyrPheValLeuHisAspValPhe	119
QY	557	CGT 559	
Db	120	Arg 120	

RESULT 5  
 Q9P8H0 PRELIMINARY: PRT 123 AA.

AC 09P8HD; 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Rph1.  
 GN Rph1.  
 OS Yarrowia lipolytica (Candida lipolytica).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes  
 OC Saccharomycetales; Dipodascaceae; Yarrowia.  
 OX NCBI\_TaxId=4952;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cordero Otero R.R.; Leplingle A.; Gallardín C.;  
 RT "RPH1 gene from Yarrowia lipolytica."  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RA Rojas Quijano R.; Leplingle A.; Gallardín C.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF260231; AAF70316.1;  
 DR HSSB; P13662; IOUN.  
 DR InterPro; IPR002075; NTF2.  
 DR Pfam; PF02136; NTF2; 1.  
 SO SEQUENCE 123 AA, 13696 MW, 356CB95C5CF0058251 CR64;

[illegible]



OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Euphorbiaceae; Oryzaeae; Oryza.  
OC NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC  
RT clone:P0415A04.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AP003345; BAB90110.1; -  
SQ SEQUENCE 146 AA; 15796 MW; 496BE9A02E1E9670 CRC64;

Alignment Scores:	
Pred. No.:	0.00295
Score:	120.00
Percent Similarity:	44.53%
Best Local Similarity:	30.47%
Query Match:	5.96%
DB:	10
Length:	146
Matches:	39
Conservative:	18
Mismatches:	59
Indels:	12
Gaps:	5

US-09-763-902B-15 (1-1146) x Q8RU32 (1-146)

```

247 188 GATCAGCGCATGTAGAGAGCTGCTAGAGAGGTGTGTCAATATTACTATAGACAATGGATATAA
248 23 ASpgInCysASpValValAlaAlaArgAlaArpHeValGluTyrTyrGlnInThrPheArpThr
249 248 .AGAAAGACGGGCACATAACCGAGGCTGTATCTGGACAAAGCCACCTTAATATGGAATGGAAAT
250 43 ASnaATGAlaAlaLeuAlaAlaLeuAlaLeuTyrGlnInThrSerMetLeuSerPheGlnGluHis
251 308 GCTGTTTCAAGGCTGGATGCCCTAAATATTTTGTGACACATGCGCTTCTAGTAGTTC
252 63 MetValAlaGluAlaGluInGluInLeuGluGluValGlyLeuLeuGluLeuProPheGlnGlnCys
253 368 CAG-----GTCAATATGTAGATTGCTCCAAACGATTCATGAGACAAACGACTAGTCCAA
254 83 ARGHISAlaValAlaCysThrValAlaAspCysGlnPro-----ThrProSerPhePro
255 422 ACTACAGTTCCTTGTGTGTGACCAAGTGGACACTGTGAAAGTTGATGGAAAACAACAACATTTCC
256 99 GtGlyGluLeuLeuValPheValSerGluAsnLeuAlaLeuAlaGluInGluInGlnLeu
257 462 ---TTCACCAACAACACTCGCTGCTGACTGCTCACTCCACTGCCACAACATTA---GtGTGG
258 119 ARGpHeSerGlnMetPhe-----GlnLeuValProAsnGlnGlnGlySerPhe
259 536 AAGATTGCAAGATGCTTCCTCGT
260 135 PheValGlnAsnAspIlePheArg

```

RESULT 7		
ID	091735	PRELIMINARY; PRT: 688 AA.
AC	091735	
DT	01-MAR-2001 (TREMBLrel. 16, Created)	
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	Rin protein.	
GN	RIN OR CG9412.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BERKELEY;	
RX	MEDLINE-20196006; PubMed-10731132;	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abill J.F., Agabryani A., An H.-J., Andrews-Fleannoch C., Baldwin D.,  
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Flitschmann W.,  
RA Foster C., Garfield A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwu C.,  
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mccarty C., Morris J., Mosnelli A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Sider-Klimos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissendach J.,  
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003701; AAG22151.1; -  
DR FlyBase; FBgn0015778; rin.  
DR InterPro; IPR002075; NTF2.  
DR InterPro; IPR005054; RNA\_rec\_mot.  
DR Pfam; PF002136; NTF2; 1.  
DR Pfam; PF00076; rrm; 1.  
DR SMART; SM00360; RRM; 1.  
DR PROSITE; PSS0102; RRM; 1.  
SQ SEQUENCE 688 AA; 74713 MW; 9E3299499E7248C19 CRC64;

Pred. No.:	0.00178
Alignment Scores:	
Score:	114.00
Percent Similarity:	39.86%
Best Local Similarity:	29.05%
Query Match:	5.66%
DB:	5
length:	688
Matches:	43
Conservative:	16
Mismatches:	63
Indels:	26
Gaps:	6

US-09-763-902B-15 (1-1146) x Q917J5 (1-688)

```

0Y      140  CTAAGAGTCCAGATGGCCAGCGTCTCGAATTTTAAACATAATGATGATCAGGCGATG  192
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      3  AATATGInSerGInGInProSerProGInSer----- 13

0Y      200  AGAGCTGCTGAGGAGATTGTGTCAATATTTACTATGAGACATGGATTAAGACAGCGGCA  258
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      14  ---ValGlyArgGlnPheValAlaArgGlnTyrTyrThrLeuLeuSerLysAlaProAsnHis  32

0Y      260  CTAACACAGCGCTATCTGTGACACAGGCGCACCTTAATA---TGAATGGAATAGCTGTTTCA  316
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      33  LeuHisArgPheTyrAlaAsnHisAsnSerSerTyrIleHisGlyLysSerLysLeuVal  52

0Y      317  GGGCGTGAATGCCCTAATAATATTTTGTGACACATGCGCTTTCATGATGTCACAGCGCAAT  376
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      53  GlyGlnAlaArgGlnIleHisAsnArgIleGln-----GlnLeuAsn  65

0Y      377  ATGTTAGATGTC-----CAACCAAGTTCATGAGCAACGAACTCAGTCCCAAACT  424
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Dd	66	PheasnaBcPcySHSLALALysLleSergInValAspAlaGlnAlaIarhLeuGlYuaNgLy	85
Qy	425	ACAGTCTCTGTGGTGACCAAGCGAAGTGTGAAGTTTGATGAGAACAACAACATTTCCTC	464
Dd	86	VALVALVALGlnValThr---GlyIuIuSerAsnAspGlyGlnPrometArGaRphE	104
Qy	485	AACAGAAACTCCGCGTCGACGCTGCAGTCCATGCCAACAAATCTGTGGGAAGATTGCA	544
Dd	105	ThgInthrPhEValLeuAlaLaGlnSerProLYslYs-----TyTYUValHIs	121
Qy	545	AGGATTTGCTTCCGTTTCCAAGAT	568
		::::	
Dd	122	AsnAspIlePhEArgTYrGlnAsp	129
 RESULT 8 Q9NH72 PRELIMINARY; PRT; 690 AA.			
ID	Q9NH72		
AC	Q9NH72:		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Rasputin.		
GN	RIN OR CG9412.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OX	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20191744; PubMed=10725247;		
RA	Pazman C., Mayes C.A., Fanto M., Haynes S.R., Mlodzik M.;		
RT	"Rasputin, the Drosophila homolog of the RasGap SH3 binding protein,		
RT	functions in Ras and Rho mediated signaling."		
RL	Development 127:1715-1725(2000).		
DR	EMBL: AF231031; AAF68949.1; -		
DR	FlyBase: FBgn0015778; rin.		
DR	InterPro: IPR002075; NTF2.		
DR	InterPro: IPR000504; RNA_rec_mot.		
DR	Pfam: PF02136; NTF2; 1.		
DR	Pfam: PF00076; rim; 1.		
DR	SMART: SM00360; RM; 1.		
DR	PROSITE: PS50102; RM; 1.		
SO	SEQUENCE 690 AA; 74985 MW; 40CSBDYE9SF5FA253 CRC64;		
 Alignment Scores:			
	Pred. No.:	0.00178	Length: 690
	Score:	114.00	Matches: 43
	Percent Similarity:	35.86%	Conservative: 16
	Best Local Similarity:	29.05%	Mismatches: 63
	Query Match:	5.66%	Indels: 26
	DB:	5	Gaps: 6
 US-09-763-902B-15 (1-1146) x Q9NH72 (1-690)			
Qy	140	GCTACAGGTCACAGATGGCACGCTCTCGATTAAAACTTAGTAGATCAGCAGCATGT	198
			:::
Dd	5	AlatrhGlnSerGAGGlnProserProGlnSer-----	15
Qy	200	AGAGTGGCTGAGAGTGTTCATAATTTACTATGACACATGATTAAGAGACGGGCA	255
Dd	16	--ValGlyArgGluPhEValArgGlnTYrTYrThLeuLeuAsnLYsAlaProAsnHis	34
Qy	260	CTAACACGAGCTATCTATCTGACAAAGGACACTTAATA---TGAATGGAATGCGTTTCA	316
Dd	35	LeuHisAspPhEArgThAsnHisnsbSerTYrIleHisGlyIuSerLYsLeuValVal	54
Qy	317	GGGCTGGATGCCCTAAATATATTTTTTGTGACACATTGCTTCTAGTAGATTCAAGGTCAAT	378
Dd	55	GlyIuNArgGlnIleHisAsnArgIleGln-----GlnLeuAsn	67
Qy	377	ATGTTAGATGCG-----CAACCAATTTCATGAGCAACCAACTCACTCCCAAACT	422

ID	Q9VFT4	PRELIMINARY:	PRT:	690 AA.
Db	68	PheasnspscysHSLsAlaIalysIIIsEsercInValaspIaIaGlnAlaIeThrlEucIyAsnGly	87	
Qy	425	ACAGTCTTGTGTGGACAGCAGTGGACAGTGTGAGTGGATGGCAAAACAACAACTTCTTC	484	
Db	88	ValValValGlnValThr---GlyGlnUeSerAsnspGlyGlnPrometargArgPhe	106	
Qy	485	AACGAACTCCGCGACACGCGACGCGACGCGACCTCCCAACATCTACTGTGTGGAGATTCGA	544	
Db	107	ThrlGlnThrPheValLeuAlaIaIaGlnSerProIysIys-----TyrTyrValHis	123	
Qy	545	AGTATGCTTCCTCCGTTTTCACAGAT	568	
Db	124	AsnAspIlePheArgTyrGlnAsp	131	
RESULT 9				
ID	Q9VFT4	PRELIMINARY:	PRT:	690 AA.
AC	Q9VFT4:			
DT	01-MAY-2000 (TREMBLrel. 13. Created)			
DT	01-MAY-2000 (TREMBLrel. 13. Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21. Last annotation update)			
DE	CG9412 protein (LDD31194p).			
GN	RIN OR CG9412.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OX	Ephydroidea; Drosophilidae; Drosophila.			
RN	NCBI_TaxID=7227;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BERKELEY;			
RX	MEDLINE=20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,			
RA	Burton G.G., Wortman J.R., Yendell M.D., Zhang Q., Chen L.X.,			
RA	Brandon R.C., Rogers Y.H.C., Blaej R.G., Champagne M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle A., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,			
RA	Abiril J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,			
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Bencs P.V., Bernan B.P., Bhandari D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,			
RA	Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke K., Davenport L.B., Davies P.,			
RA	de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,			
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,			
RA	Glock A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,			
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwu C.,			
RA	Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,			
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,			
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,			
RA	Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,			
RA	Spieler E., Spirdling A.C., Stapleton M., Strong R., Sun E.,			
RA	Svirskas R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,			
RA	Williams S., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,			
RA	Ye J., Yen R.F., Zaerli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,			
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,			
RT	"The genome sequence of Drosophila melanogaster."			
RL	Science 287:2185-2195(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BERKELEY;			

1

RA. Tortumt M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,





RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R;  
RT "Direct Submission";  
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases  
DR EMBL; AC006605; AAK85441.2; -  
DR InterPro; IPR000357; HEAT\_repeat.  
DR PROSITE; PS00077; HEAT\_REPEAT; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 1378 AA; 153109 MW; E2DA7981417FD1CB CRC64

```

Alignment Scores:
Pred. No.:      0.249      Length:      1378
Score:          95.00      Matches:      36
Percent Similarity: 38.61%  Conservative: 25
Best Local Similarity: 22.78% Mismatches: 75
Query Match:    4.72%      Indels:      22
DB:             5          Gaps:         5
US-09-763-902B-15 (1-1146) x Q95YF0 (1-1378)

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:24:08 ; Search time 17.3988 Seconds  
(without alignments)  
5463.815 Million cell updates/sec

Title: US-09-763-902b-15  
Perfect score: 2013  
Sequence: 1-tacttcgggagagagatg99.....tttaagtaattttgtttt 1146

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 112892 segs, 41476328 residues  
Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ .n2p.model -DEV-xlp  
-Q-/cg92.1/USPTO.spool/US0976302/runal\_24062003\_130340\_12418/app\_query.fasta.1.2254  
-DB-Swissprot.40 -QEMT-fastan -SUFFIX-n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE-LOCAL -OUTMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR-US0976302.ecgn.1.1.46-ecgnal\_24062003\_130340\_12418 -NCPD=6 -ICPD=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV-TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database: SwissProt.40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	749	37.2	142	1	NXT2_HUMAN
2	586	29.1	140	1	NXT1_MOUSE
3	584	29.0	140	1	NXT1_HUMAN
4	310	15.4	133	1	NXT1_DROME
5	207.5	10.3	137	1	NXT1_CAEEL
6	124	6.2	132	1	NFT2_ORYSA
7	115	5.7	127	1	NFT2_HUMAN
8	113	5.6	123	1	NFT2_SCHPO
9	112.5	5.6	126	1	NFT2_ARATH
10	108.5	5.4	124	1	NFT2_NEUCR
11	108	5.4	127	1	NFT2_XENTL
12	104.5	5.2	124	1	NFT2_CANAL
13	101.5	5.0	125	1	NFT2_YEAST
14	99.5	4.9	133	1	NFT2_CAEEL
15	88.5	4.4	482	1	G3B2_HUMAN
16	88.5	4.4	482	1	G3B2_MOUSE
17	87	4.3	643	1	NOS2_ALCEU
18	85	4.2	1452	1	VGL2_FITPV

C	19	81	4.0	358	1	VEGD_MOUSE	P97946 mus musculu
C	20	81	4.0	515	1	YNBT_YEAST	P53741 saccharomyc
C	21	81	4.0	579	1	UL25_VZVD	P09287 varicella-z
C	22	80.5	4.0	2151	1	RRPL_HANTV	P23456 hantaa vir
C	23	79	3.9	260	1	SVS1_YEAST	Q12254 saccharomyc
C	24	78.5	3.9	450	1	ACHX_HUMAN	Q99226 homo sapien
C	25	78.5	3.9	1176	1	SLAP_BACSH	P74293 bacillus sp
C	26	78	3.9	287	1	FFG_SYNY3	P74293 bacillus sp
C	27	77.5	3.8	309	1	YDAA_HAEIN	P44195 haemophilus
C	28	77.5	3.8	530	1	UDBH_HUMAN	Q75795 homo sapien
C	29	77.5	3.8	776	1	ARHG_HUMAN	Q15052 homo sapien
C	30	77.5	3.8	910	1	XO68_HUMAN	P34607 caenorhabdi
C	31	77.5	3.8	1451	1	WGL2_CVCAI	P36300 canine ente
C	32	77.5	3.8	2151	1	RRPL_SEOUB	P27314 seoul. virus
C	33	77	3.8	778	1	PRTY_SCHPO	Q42933 schizosacch
C	34	76.5	3.8	355	1	S1B8_MOUSE	Q64687 mus musculu
C	35	76.5	3.8	434	1	G3BP_HUMAN	Q94260 schizosacch
C	36	76.5	3.8	4128	1	PRKD_HUMAN	Q43915 homo sapien
C	37	76	3.8	740	1	CATE_STRRE	P08784 streptomyce
C	38	76	3.8	901	1	SOK1_YEAST	P40317 saccharomyc
C	39	75.5	3.7	626	1	RPOC_SYNY3	P74177 synechocyst
C	40	75.5	3.8	753	1	CKAA_BACUF	O32321 bacillus th
C	41	75	3.7	594	1	VEGD_HUMAN	Q43915 homo sapien
C	42	75	3.7	594	1	KR2_HSVB	P28966 equine herp
C	43	74	3.7	375	1	YKUS_CAEEL	Q17778 caenorhabdi
C	44	74	3.7	1112	1	RA13_SCHPO	P28706 schizosacch
C	45	73.5	3.6	363	1	VP43_NPVAC	P34050 autographa

## ALIGNMENTS

RESULT 1	NXT2_HUMAN	STANDARD;	PRT; 142 AA.
ID	NXT2_HUMAN		
AC	Q9NPJ8; Q9H8U0; Q9NRL7; Q9Y3M4; Q9Y3M5; Q9NQ64;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	NFT2-related export protein 2 (p15-2 protein) (DC9) (BM025).		
GN	NXT2.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI-TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).		
RC	TISSUE=Testis;		
RX	MEDLINE=20528640; PubMed=11073998;		
RA	Herold A., Suyama M., Rodrigues J.P., Braun I.C., Kutay U.,		
RA	Carmona-Fonseca C., Bork P., Izaurralde E.,		
RT	"Tap (NXT1) belongs to a multigene family of putative RNA export		
RT	factors with a conserved modular architecture."		
RL	Mol. Cell. Biol. 23:8996-9008(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RA	Kang Y., Cullen B.,		
RT	"p15-2, a homologous protein of p15, interacts with Tap."		
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RA	Li Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.,		
RT	"Novel genes expressed in human dendritic cell."		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RC	TISSUE=Ovarian carcinoma;		
RA	Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,		
RA	Wagatsuna M., Hosokiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,		
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,		
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,		
RA	Yamamoto T., Wakamatsu A., Nakamura Y., Nagaheri K., Masuno Y.,		
RA	Ninomiya K., Iwayanagi T.,		











CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AB011262; BAA81910.1;  
DR HSSP: P13662; 1A2K.  
DR InterPro: IPR002075; NTF2.  
DR Pfam: PF02136; NTF2.1.  
DR PROSITE: PS50177; NTF2-DOMAIN; 1.  
DR TRANSPORT: Protein transport.  
KW DOMAIN  
FT DOMAIN 6 NTF2.  
SQ SEQUENCE 122 AA; 13354 MW; 84100EBB62AB1BD1 CRC64;

Alignment Scores:  
Pred. No.: 3.96e-05 Length: 122  
Score: 124.00 Matches: 39  
Percent Similarity: 47.54% Conservative: 19  
Best Local Similarity: 31.97% Mismatches: 50  
Query Match: 6.16% Indels: 14  
DB: 1 Gaps: 5

US-09-763-902b-15 (1-1146) x NTF2\_ORFSA (1-122)

QY 206 GCTGAGGAGTTTTCATATTTACTATGACACATGATAAAGACGGCAGTAAACC 265  
DB 7 AAlaylaAlpheValaIgUHSlyTyrRyRhrhphesprhAaAgpProAlaLeuVal 26  
QY 266 AGCGTGTATGTGCAAGACCCCTTAATATGAGTAAGTAAGTCTGTTTCAGGCTGAT 325  
DB 27 SerLeuYrGlnAspGlySerMetLeuThrPheGlnGlnInpHeuGlyAlaAla 46  
QY 326 GCCCTAATATATTTTGTGACACATTCCTTCAGTGAAG-----TTCAGAGTAAATATG 379  
DB 47 AAlaLeaGlyLysLeuGlySerLeuProPheAlaGlnCysHsHsAspIleAsnThr 66  
QY 380 TTAGATGACCAACGATTCATGACGACGACGATGCTCCCACTACAGTCTGTTGTG 439  
DB 67 ValAspGlyGlnPro-----SerIlyProGlnGlyGlyMetLeuValPhe 81  
QY 440 ACCAGTGAAGTGTGAGTTGATGGAACAAACAGAT-----TTCCTCACCAGAAC 493  
DB 82 ValSerGlySerLeuArg---ThrGlyProAspGlnHsProLeuLysPheSerGlnMet 100  
QY 494 TTCCTGCTGACTGCTCAGTCCAGTCCCAACAAATCTGTGGAAGATTCAGATGATGC 553  
DB 101 Phe-----GlnLeuLeuProAlaGlyLysAspPheValGlnAsnAspMet 116  
QY 554 TTCCTG 559  
DB 117 PheArg 118

RESULT 7  
NTF2\_HUMAN STANDARD; PRT: 127 AA.  
ID NTF2\_HUMAN  
AC P13662;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE Nuclear transport factor 2 (NTF-2) (Placental protein 15) (Pp15).  
GN NTF2 OR Pp15.  
OS Homo sapiens (Human),  
OS Mus musculus (Mouse), and  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606, 10090, 10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;

RX MEDLINE=88247772; PubMed=3380696;  
RA Grundmann U., Nerlich C., Rein T., Lottspeich F., Knepper H.A.;  
RT "Isolation of cDNA coding for the placental protein 15 (Pp15).";  
RL Nucleic Acids Res. 16:4721-4721(1988).  
RN [2]  
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RC SPECIES=Human;  
RX MEDLINE=95263674; PubMed=7744965;  
RA Paschal B.M., Gerace L.;  
RT "Identification of NTF2, a cytosolic factor for nuclear import that  
RT interacts with nuclear pore complex protein p62.";  
RL J. Cell Biol. 129:923-937(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human, and Mouse;  
RA Strausberg R.;  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Mouse; STRAIN=C57BL/6J; TISSUE=Kidney;  
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A.,  
RA Arakawa T., Carrincci P., Fukuda S., Fukunishi Y., Furuno M.,  
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hirazaki T., Hori F.,  
RA Inotani K., Ishii Y., Itoh M., Izawa M., Kato H., Kawai J., Kojima Y.,  
RA Kono H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A.,  
RA Nishi K., Nomura K., Numazaki R., Ono M., Okazaki Y., Okido T.,  
RA Owa C., Saito H., Saito R., Sakai C., Sakai K., Sano H., Sasaki D.,  
RA Shubata K., Shibata Y., Shingawa A., Shiraki T., Sogabe Y.,  
RA Suuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y.,  
RA Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
RA Muramatsu M., Hayashizaki Y.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RA Kent H.M., Clarkson W.D., Bullock T.L., Stewart M.;  
RT Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
RC SPECIES=Human;  
RX MEDLINE=96300274; PubMed=8757804;  
RA Bullock T.L., Clarkson W.D., Kent H.M., Stewart M.;  
RT "The 1.6-A resolution crystal structure of nuclear transport factor 2  
RT (NTF2).";  
RL J. Mol. Biol. 260:422-431(1996).  
RN [7]  
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
RX MEDLINE=98035028; PubMed=9368653;  
RA Clarkson W.D., Corbett A.H., Paschal B.M., Kent H.M., McCoy A.J.,  
RA Gerace L., Silver P.A., Stewart M.;  
RT "Nuclear protein import is decreased by engineered mutants of nuclear  
RT transport factor 2 (NTF2) that do not bind GDP-Ran.";  
RL J. Mol. Biol. 272:716-730(1997).  
RN [8]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF COMPLEX WITH RAN.  
RC SPECIES=Rat;  
RX MEDLINE=98202576; PubMed=9533885;  
RA Stewart M., Kent H.M., McCoy A.J.;  
RT "Structural basis for molecular recognition between nuclear transport  
RT factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase  
RT Ran.";  
RL J. Mol. Biol. 277:635-646(1998).  
RN [9]  
RP FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS.  
CC INTERACTS WITH THE NUCLEOPORIN P62 AND WITH RAN. ACTS AT A  
CC RELATIVELY LATE STAGE OF NUCLEAR PROTEIN IMPORT, SUBSEQUENT TO THE  
CC INITIAL DOCKING OF NUCLEAR IMPORT LIGAND AT THE NUCLEAR ENVELOPE.  
CC COULD BE PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT  
CC ASSEMBLE AT THE PORE COMPLEX DURING NUCLEAR IMPORT.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

US-09-763-902B-15 (1-1146) x NTF2-SCHEPO (1-123)

Oy 206 GCTGAGAGAGTTTGCATATTTACTATGAGACAATGATTAAGAAGACGGGCACTAAC 265  
||| :||| |||||:|||| ||| ||| :|||:  
Db 8 AlaThrGlnPheThrGlnPheTyrTyrGlnThrPheAspSerAspArgSerGlnLeuSer 27

```
CC or send an email to licensed@slsb.ch).
DR EMBL AC069471: AAC51491.1; -.
DR HSSP: P13662; IOUN.
DR InterPro: IPR002075; NTF2.
DR Pfam: PF02136; NTF2_1.
DR PROSITE: PS50177; NTF2_DOMAIN; 1.
KW Transport; Protein transport.
KM DOMAIN
SQ SEQUENCE 126 AA; 14002 MW; 263E06A7AB903E3B CRC64;
FT 9 NTF2.
Alignment Scores:
Pred. No.: 0.000687 Length: 126
Score: 112.50 Matches: 33
Percent Similarity: 47.93% Conserved: 25
Best local Similarity: 27.27% Mismatches: 52
Query Match: 5.59% Indels: 11
DB: 1 Gaps:. 5
US-09-763-902B-15 (1-1146) x NTF2_ARATH (1-126)
OY 206 GCTGAGAGATTGTGCATTAATTACTATGACAGCAATGGATAAAGACGGCAGTAAACC 265
Db : ::::: ||||| ||||| || | || | || |
10 SerLyshAlpheValGlUnHstYrTSerThrPhenSPThrAsnArgValGIyleUla 29
OY 266 AGCGTCGTATCTGGACAAGGCCCATTTAATATGGAATGGAATGCGTITTCAGGCGTGAT 325
Db : ||||| ::::: || | || | || |
30 GlyLeuTyClngInlunIasemetleuthrPhelngInglYnIlsgInglYalGln 49
OY 326 GCCCTAAATATATTTTTTGACACATTTGCT-----TCTAGTAGCTTCCAGGTCATATATG 379
Db : ::::: ||||| ||::: ||:::
50 SerIleValAlaLysLeuThrSerLeuProPhelngInclncYSylShIsHisIleSerThr 69
OY 380 TTGATGATGCCAACCACTTCATGACGACCACTAGTCCCAACAGTACAGTCTTGTTGTG 439
Db : ::||| ||||| ||::: ||:::
70 ValAspCysGlnPro-----SergLYPrIoAlaSerGlyMetLeuValPhe 84
OY 440 ACCAGTGAAGCACTGTGAGTTTGATGGAACAACAACAATTTTC--TTCAACAGAACTTC 498
Db : ||||| ::::: || | ::::: || |
85 ValSerClYsnIeueGlnIeuAlaglYgInglUnHsAlaleuLYrsPheSerGlnMetPhe 104
OY 497 CTGCTGACGCTGCTAGTGCACCTCCCAACAATACTGTGTGTGAGAAGATGCAAGATGCTGCTC 555
Db : || | ||||| ||::: ||::: || |
105 HisLeu-----MetProThrProGlnGlySerPheTY---ValPheAsnAspIlePhe 121
OY 557 CGT 559
Db : || |
122 Arg 122
RESULT 10
NTF2_NEUCR
ID NTF2_NEUCR STANDARD; PRT; 124 AA.
AC P81102;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear transport factor 2 (NTF-2).
GN NTF-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezilomycelia; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_Taxid=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BD-A / FGSC 1858;
RA Meyer U., Meyer W., Tschel D., Toeken K., Rensing L.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
-i FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE
PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).
CC -1 SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1 SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
```







```
Db      |||||.....|||          ||::: |||
        2 SerLeuAspPheHisThr-----LeuAlaGlnInsnpHeThrgIn 14

OY      224 ATTTACTATGAGACAATGGATATAAAGAACGGGCACCTAACAGCGGTGATTCGACAAAG 283
         |||||   |||   |||   |||   |||   |||   |||   |||
Db       15 PheTyTYrTAsnGlnPheaspPrHrsApArgSerGlnLeuGLyAsnLeuTyTrtAsnGln 34
         :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

OY      284 GCCACCTTAATATATGAATGGAATGGAATGCTGTTTGACGGCGTGGATGCCCTAAATATATTTTTT 343
         ::::  |||   ::::  ::::  ::::  ::::  |||   ::::
Db       35. SerMetLeuThrPheGlnIuThrSerGlnLeuGlnGlyAlaLysAspLLevalGluTylsLeu 54
         ::::  |||   ::::  ::::  ::::  ::::  ::::  ::::

OY      344 GACACATTGCCCTTCAGTAGAGTCCAG-----GTCAATATGTTAGATTGCCAACAGTT 397
         ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       55 ValSerLeuProPheGlnLysValGlnHisArgIleThrThreusAspLagInProAla 74

OY      398 CATAGGCAAGCAACCTCAGTCCCACAACTACAGTTCCTGTGTGTGACCAGCTGGAAG 457
         |||||.....|||          |||||.....|||
Db       75 SerProAsnGly-----AspValIleValMetIleThrGlyAspLeuLeu 89

OY      458 TTGTGAT-----GGAACAACAACACATTCTCTCAACCAACAACATCTCCTGTGATGCTCAG 511
         |||   |||   |||   |||   |||   |||   |||   |||
Db       90 IleAspGlnGlnGlnAsnProGlnArg---PheSerGlnValPheHisLeu----- 105

OY      512 TCCACTCCCAACAATACGTGTGTGAGAGATTCGAAGGATGCTCGCT 559
         ||:::  ::::  ::::  ::::  ::::  ::::  ::::  ::::
Db       106 ---IleProAspGlyAsnSerTyTrtValPheAsnAspIlePheArg 120
```

RC TISSUE-Brain;  
RA Kennedy D., Ru K., Matlick J.S.;  
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE-B-cell;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
RN [6]  
RP INTERACTION WITH IKAPPABALPHA.  
RX MEDLINE-20549669; PubMed-10969074;  
RA Prigent M., Barlat I., Langen H., Dargemont C.;  
RT "Ikappabalpha and Ikappabalpha/NF-kappa B complexes are retained in  
RT the cytoplasm through interaction with a novel partner, RasGAP  
RT SH3-binding protein 2".  
RL J. Biol. Chem. 275:36441-36449(2000)  
CC -1- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA  
CC TRANSPORT (POTENTIAL).  
CC -1- SUBUNIT: BINDS TO THE N-TERMINAL DOMAIN OF IKAPPABALPHA.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A (SHOWN HERE) AND B; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF145284; AAD51932.1; -  
DR EMBL: AB014560; BA31635.1; -  
DR EMBL: AF051311; AAC15705.1; -  
DR EMBL: AF053535; AAC95292.1; -  
DR EMBL: BC011731; AAH11731.1; -  
DR HSSP: P09651; 1HA1.  
DR InterPro: IPR002075; NTF2.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR Pfam: PF00076; rrm; 1.  
DR SMART: SM00360; RRM; 1.  
DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
DR PROSITE: PS50102; RRM\_1; FALSE\_NEG.  
DR PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
KW Transport; RNA-binding; Alternative splicing.  
FT DOMAIN 11 133 NTF2.  
FT DOMAIN 331 409 RNA-BINDING (RRM).  
FT DOMAIN 134 223 GUU-RICH.  
FT DOMAIN 419 479 GLY-RICH.  
FT VARSPPLIC 243 275 MISSING (IN ISOFORM B).  
FT CONFLICT 267 267 S -> P (IN REF. 2).  
FT CONFLICT 359 359 E -> V (IN REF. 3).  
FT CONFLICT 460 460 M -> I (IN REF. 3).  
SQ SEQUENCE 482 AA: 54111 MW: 25456CA3F1AAE218 CRC64;

## Alignment Scores:

Pred. No.: 0.342 Length: 482  
Score: 88.50 Matches: 34  
Percent Similarity: 42.40% Conservative: 19  
Best Local Similarity: 27.20% Mismatches: 63  
Query Match: 4.40% Indels: 9  
DB: 1 Gaps: 4

US-09-763-902B-15 (1-1146) x G3B2\_HUMAN (1-482)

QY 212 GAGTTTCAATATTACTAGACATGATAAAGAGAGCGGCACTATACGCGTG 271  
|||||  
DB 14 Gtuphevalnrglnttyttrhrlenuasnlysalaprogltutrylenuhlsarqph 33  
272 TATCTGACAAAGGCCACTTATATGATGAAATGGAATGCTGTTTCAGCG----- 319

DB 34 TyrGIArgaAnSerSertYrValhISGLyValaAsPaLaSerGIlySProGInGu 53  
|||  
QY 320 ---CTGCATGCCCTAAATATTTTTCACACATTGCTTACGTAGTCCAGTCAAT 376  
|||  
DB 54 AlaValTYrGLyGlnAsnAspIleHIShISLysValLeuSerLeuAsnPheserGIuys 73  
|||  
QY 377 ATGTTAGATTCCCAACCACTTCATGAGCAAGCACTCAGTCCCAACCTACAGTCTTGT 436  
|||  
DB 74 HIsThrIySLeArgHISValaSPalaHISAlaThrLeuSerAsp---GLyValVal 92  
|||  
QY 437 GTGACCACTGGAAGTGTGAAGTTGATGAGAAACAACACATTCTTCAACCACTTC 496  
|||  
DB 93 GlnValMetGLyLeuLeuSerAsnSerGIyGlnProGluArgLysPhemetGlnThrPhe 112  
|||  
QY 497 CTGCTGACTGCTCAG---TCCACTCCCAACATACTGTGTGGAAGATTGCAAGTGTTC 553  
|||  
DB 113 ValLeuAlaProGluGlySerValProAsnLys-----PhetYrValhISAsnAspMet 130  
|||  
QY 554 TTCGCTTTCAGAT 568  
|||  
DB 131 PheaTgTYrGIuAsp 135

Search completed: June 24, 2003, 19:32:13  
Job time : 21.398 secs

OY	21	TTTGCAATATTACATGATAGACAATGCAATAAAGAAGCGGCACTAACCAAGCGGTAT	27
OY			
Db	14	PheIleGlnHisIytrIyGlnLeuPheAspAsnAphrIyGlnLeuGlyAlaIleIytr	33
OY	275	CTGGACAGACGCCACTTAATATGAAATGGAAATGCGTTCACGGCTGATGCCCTTAAT	33
OY			
Db	34	IleAspAlaSerGlyLeuThrIyTrpGluGlnGlnPheGlnGlyAlaAlaIleVal	53
OY	335	AATTTTGTGACACATTCGCTTCTATGTAGTCCAG----	388
OY			
Db	54	GlnIlyLeuSerSerLeuProPheGlnIlyIleGlnHisSerIleThrAlaGlnAspHis	73
OY			

QY 389 CACACATTCATGAGCAAGCAACTACGCCAATCAGTCTTGTGTGACCACTGGA 448  
||||| : : : : :  
Db GlnProThrProAsp-----SerCysIleIleSerMetValGly 87  
QY 449 ACTGTGAAGTTTGTGGAACAACAACATTTCTTCAACCAAGACTTCGTGACTGCT 508  
||||| : : : : :  
Db GlnLeuLysAlaAspGlnAspProIleMetClyPheHisGlnMetPheLeu----- 105  
QY 509 CAGCCACACTCCCAACATACGTGTGAGAGTTGCAAGTTCAGTCTCCCGT 559  
Db 106 -----LysAsnIleAsnAspAlaTrpValCysThrAsnAspMetPheArg 120  
RESULT 2  
T38039  
Probable nuclear transport factor 2 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 15-Sep-2000  
C:Accession: T38039; T37728  
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Whitehead, S.; Churcher, C.M.  
submitted to the EMBL Data Library, August 1999  
A:Reference number: Z21764  
A:Accession: T38039  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-123 <MCD>  
A:Cross-references: EMBL:AL109951; PIDN:GAB53052.1; GSPDB:GN00066; SPDB:SPAC1B9.01c  
A:Experimental source: strain 972h; cosmid C1B9  
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z21740  
A:Accession: T37728  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: SL\_77-123 <DEV>  
A:Cross-references: EMBL:Z68197; PIDN:CAA92380.1; GSPDB:GN00066; SPDB:SPAC1SF9.03c  
C:Genetics:  
A:Gene: SPDB:SPAC1B9.01c  
A:Map position: 1  
A:Introns: 3/1, 30/3; 54/3; 88/3  
C:Superfamily: yeast nuclear transport factor NTF2  
Alignment Scores:  
Pred. No.: 0.00124 Length: 123  
Score: 113.00 Matches: 34  
Percent Similarity: 47.93% Conservative: 24  
Best Local Similarity: 28.10% Mismatches: 51  
Query Match: 5.61% Indels: 12  
DB: 2 Gaps: 4  
US-09-763-902B-15 (1-1146) x T38039 (1-123)  
QY 206 GCTGAGAGTTTGTCAATATTACTATGAGCAATGATAAAGAGCGGCACTAAC 265  
||||| : : : : :  
Db AAlaTrGlnPheThrGlnPheTyrTyrGlnThrPheAspSerArgSerGlnLeuSer 27  
QY 266 AGGCTGTATCTGACCAAGGCCACCTTAATAGATGAATGCTGTTCAAGGCTGAT 325  
||||| : : : : :  
Db 28 SerLeuTyrArgGlnGlnSerMetCysLeuSerPheGlnGlnLeuGlnGlyThrLys 47  
QY 326 GCCCTAATAATTTTTTTGACACATTCGCTTCTAGTGCAG-----GTCAATATG 379  
||||| : : : : :  
Db 48 AlaIleValGlnLysLeuValSerLeuProPheGlnArgValGlnHisArgIleSerThr 67  
QY 380 TTAATATGCCAACCAAGTTCATGAGCAAGCAACTAGTCCCAACTACAGTCTGTTGTG 439  
||||| : : : : :  
Db 68 LeuAspAlaGlnPro-----ThrGlyThrThrGlySerValIleValMet 82  
QY 440 ACCAGTGAAGTGTGAGATTGATGAGAAACAACAA-----CATTTCTTCAACGAGACTTC 496  
||||| : : : : :  
Db 83 ValThrGlyGlnLeuLeuLeuAspGlnGlnMetAlaGlnArgTyrSerGlnValPhe 102  
QY 497 CTGCTGACTGCTCAGTCACATCCCAACAATACGTGTGTGGAAGATTCAGTATGCTTC 556

Db 103 HisLeu-----ValAsnAsnAsnGlyAsnTyrTyrValLeuAsnAspLeuPhe 118  
QY 557 CGT 559  
|||  
Db 119 Arg 119  
RESULT 3  
B86405  
Probable nuclear transport factor 2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: B86405  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federer, N.A.; Kaul, S.; White, O.; Alonso, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talloker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: B86405  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-126 <STO>  
A:Cross-references: GB:AE005172; NID:911024877; PIDN:AAG26961.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
C:Superfamily: yeast nuclear transport factor NTF2  
Alignment Scores:  
Pred. No.: 0.00141 Length: 126  
Score: 112.50 Matches: 33  
Percent Similarity: 47.93% Conservative: 25  
Best Local Similarity: 27.27% Mismatches: 52  
Query Match: 5.59% Indels: 11  
DB: 2 Gaps: 5  
US-09-763-902B-15 (1-1146) x B86405 (1-126)  
QY 206 GCTGAGAGTTTGTCAATATTACTATGAGCAATGATAAAGAGCGGCACTAAC 265  
||||| : : : : :  
Db 10 SerLysAlaPheValGlnHisTyrTyrSerThrPheAspThrAsnArgValGlyLeuAla 29  
QY 266 AGGCTGTATCTGACCAAGGCCACCTTAATAGATGAATGCTGTTCAAGGCTGAT 325  
||||| : : : : :  
Db 30 GlyLeuTyrGlnGlnAlaSerMetLeuThrPheGlnGlnGlnLysIleGlnGlyValGln 49  
QY 326 GCCCTAATAATTTTTTTGACACATTCGCT-----TTAGTGAAGTCCAGTCAATATG 379  
||||| : : : : :  
Db 50 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnCysLysHisIleSerThr 69  
QY 380 TTAGATGCCAACCAAGTTCATGAGCAAGCAACTAGTCCCAACTACAGTCTTGTG 439  
||||| : : : : :  
Db 70 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 84  
QY 440 ACCAGTGAAGTGTGAGATTGATGAGAAACAACAACATTC-----TTCAACGAGACTTC 496  
||||| : : : : :  
Db 85 ValSerGlyAsnLeuGlnLeuAlaGlyLysIleHisAlaLeuLysPheSerGlnMetPhe 104  
QY 497 CTGCTGACTGCTCAGTCACATCCCAACAATACGTGTGTGGAAGATTCAGTATGCTTC 556  
||||| : : : : :  
Db 105 HisLeu-----MetProThrProGlnGlySerPheTyr---ValPheAsnAspIlePhe 121  
QY 557 CGT 559  
|||  
Db 122 Arg 122  
RESULT 4  
H86248

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H86398  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-122 <STO>  
A:CROSS-references: GB:AE005172; NID:g9802547; PIDN:AAPF9749.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F17L21.10  
A:Map position: 1  
C:Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:
0.0091	105.00	45.90%	29.51%
Length: 122	Matches: 36	Conservative: 20	Mismatches: 52
Gaps: 5	Indels: 14		

US-09-763-902B-15 (1-1146) x H86398 (1-122)

OY 206 GCCTGAGAGATTTTGCATTATTTACTATGACCAATGGATAAAGAAGCGGCACTAAC 265  
Db ||||| ||||| ||||| ||||| |||||  
OY 7 ATATyAlaphavalGluHISTYrTThPheaspAlasnaPrroGIyleVal 26  
Db ||||| ||||| ||||| ||||| |||||  
OY 266 AGCGTGATGTGACAAGCCACCCTTAATGTGAATGAATGCTGTTTCAGGCGTGAT 335  
Db ||||| ||||| ||||| ||||| |||||  
OY 27 SerleutylngIngluyserMetLeuthPhedIncluylnlysllegInglyserGln 46  
Db ||||| ||||| ||||| ||||| |||||  
OY 326 GCCCTAAATATTTTTTTTGACACATTCCTTCTTAGTACGTTCCAGGTCATATG----- 379  
Db :::: ||||| ||||| ||||| |||||  
OY 47 AsnIleValAlalysLeuthrgIlyLeuprophedInclnglyshISAsnIlethrThr 66  
Db ||||| ||||| ||||| ||||| |||||  
OY 380 TTGATATGCCAACAGTCAATGACGACCAAGCAAGCAAGTCCAAACTACAGTCTGTGTG 439  
Db ||||| ||||| ||||| ||||| |||||  
OY 67 ValnsPyssgInPro-----SerIlyProAlaglylMetLeuValPhe 81  
Db ||||| ||||| ||||| ||||| |||||  
OY 440 ACCAGTGGAACTGTGATGATGAACAACAAACAACTTTC-----TTCACACGAAC 493  
Db ||||| ||||| ||||| ||||| |||||  
OY 82 ValSerItylsnleugInleuAlaglY---GlucInHisAlaleuLyIsPheSerGlnMet 100  
Db ||||| ||||| ||||| ||||| |||||  
OY 494 TTCCTGTGACTGCTCAGTCCACATCCCAACAAATCTGTGTGAAGATTGCAAGTGATGC 553  
Db ||||| ||||| ||||| ||||| |||||  
OY 101 PheHisLeuIleSer-----AsnGlnGlyasnTryrTyValPheAsnAspIle 116  
Db ||||| ||||| ||||| ||||| |||||  
OY 554 TTCGCT 559  
Db ||||| ||||| ||||| ||||| |||||  
Db 117 PheArg 118

RESULT 6  
S50467  
nuclear transport factor NTF2 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YER009w  
C:Species: Saccharomyces cerevisiae  
C>Date: 28-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 24-Sep-1999  
C:Accession: S50467; S41793; S72237  
R:Dietrich, F.S.  
submitted to the EMBL Data Library, December 1994  
A:Description: The sequence of S. cerevisiae cosmid 9537, 9581, 9495, 9867, and lamb  
A:Reference number: S50459  
A:Accession: S50467  
A:Molecule type: DNA  
A:Residues: 1-125 <DIE>  
A:CROSS-references: EMBL:U18778; NID:g603592; PIDN:AAB64542.1; PID:g603601; MIPS:YERO  
R:Haarer, B.K.; Petzold, A.S.; Brown, S.S.  
submitted to the EMBL Data Library, July 1993  
A:Description: Identification of mutations that are synthetically lethal with altered  
A:Reference number: S41793  
A:Accession: S41793  
A:Molecule type: DNA  
A:Residues: 1-84 <HAA>  
A:CROSS-references: EMBL:L22204; NID:g3447714; PIDN:AAB49379.1; PID:g3447715

R.Haerter, B.K.: Corbett, A.; Kweon, Y.; Petzold, A.S.; Silver, P.; Brown, S.S.  
Genetics 144, 495-510, 1996

A:Title: SEC3 mutations are synthetically lethal with profilin mutations and cause defec  
A:Reference number: S72237; MUID:970444444; PMID:8889515

A:Accession: S72237

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-84 <HMA>

A:Cross-references: EMBL:L22204; NID:G347714; PIDN:AA849379.1; PID:G347715

C:Genetics:

A:Gene: SGD:NIF2

A:Cross-references: SGD:S0000811; MIPS:YER009\*

A:Map position: 5R

C:Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:

Pred. No.:	0.0219	Length:	125
Score:	101.50	Matches:	38
Percent Similarity:	41.91%	Conservative:	19
Best Local Similarity:	27.94%	Mismatches:	58
Query Match:	5.04%	Indels:	21
DB:	2	Gaps:	6

US-09-763-902B-15 (1-1146) x S50467 (1-125)

```

OY 164 TCTGTGATTTTAAACTTATGTAGATCAGCAGTACGCTGAGAGTTTGTCAAT 223
    |||||
DB 2 SerLeuAspPheAsnThr-----LeuAlaGlnAsnPheThrGln 14
OY 224 ATTACTGTGAGACATGATTAAGAGACGGCCTAACCGCTTATCTGGACAG 283
    |||||
DB 15 PheYrYrYrAsnGlnPheAspThrAspArgSerGlnLeuGlnAsnLeuYrYrAsnGln 34
OY 284 GCCACCTTAATATGGAATGGAATGCTGCTTACAGGCGCGATGCCCTAAATATTTT 343
    |||||
DB 35 SerMetLeuThrPheGlnThrSerGlnLeuGlnGlnAlaLysAspPheValGlnLysLeu 54
OY 344 GACACATTCCTTACAGTACAGTCCAG-----GTCAATATGTGTGATGACCAACACTT 397
    |||||
DB 55 ValSerLeuProPheGlnLysValGlnHisArgIleThrThrLeuAspAlaGlnProAla 74
OY 398 CATGAGCAGACACTCACTCCCAACTACAGTCTTGTGTGACACAGTGAACCTGTGAAG 457
    |||||
DB 75 SerProAsnGly-----AspValLeuValMetIleThrGlyAspLeuLeu 89
OY 458 TTGTAT-----GGAAGAACACATTTCTTCAACCGCAACTTCTCTGACTGCTCAG 511
    |||||
DB 90 IleAspGlnGlnGlnAsnProGlnArg---PheSerGlnValPheHisLeu----- 105
OY 512 TCACATCCCAACAATACATGCTGTGAGAGATGCAAGTGAATGCTGCTCGT 559
    |||||
DB 106 ---IleProAspGlyAsnSerYrYrYrValPheAsnAspIlePheArg 120

```

# RESULT 7

hypothetical protein R05D11.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T23921

R:White, S.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19818

A:Accession: T23921

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-133 <WII>

A:Cross-references: EMBL:Z75546; PIDN:CAA93890.1; GSPDB:GN00019; CESP:R05D11.3

A:Experimental source: clone R05D11

C:Genetics:

A:Gene: CESP:R05D11.3

A:Map position: 1

A:introns: 33/1; 63/3; 96/3

C:Superfamily: yeast nuclear transport factor NTF2

## Alignment Scores:

Pred. No.:	0.0365	Length:	133
Score:	99.50	Matches:	34
Percent Similarity:	48.03%	Conservative:	27
Best Local Similarity:	26.77%	Mismatches:	47
Query Match:	4.94%	Indels:	19
DB:	2	Gaps:	6

US-09-763-902B-15 (1-1146) x T23921 (1-133)

```

OY 206 GCTGAGAGGTTTGCATATTTACTATAGACATGAT-----AAAGA 250
    |||||
DB 11 AlaYsAlaPheIleGlnHisIstYrYrSerLysPheAspValGlyAspGlyMetSerArg 30
OY 251 AGACGGCCTAACACAGCCTGTAT---CTGACAAGCGCACCTTAATATGAATGAAT 307
    |||||
DB 31 AlaGlnGlyLeuSerAspLeuYrYrAspProGlnAsnSerYrMetThrPheGlnGlyGln 50
OY 308 GCTGTTTCAGGCGTGGATGCCCTAAATATTTTGTGACACATTCCTTCTAGTCAATTC 367
    |||||
DB 51 GlnAlaLysGlyArgAspGlyIleLeuGlnLysPheThrThrLeuGlnPheThrLysIle 70
OY 368 CAG-----GTCAATATGTGTGATGACCAACAGTTCATGACCAACACTCAGTCCCA 421
    |||||
DB 71 GlnArgAlaIleThrValIleAspSerGlnProLeuYrYrAspGlySerIleGln----- 88
OY 422 ACTACAGTTCCTTGTGCTGACACAGTGAACCTGATGATGGAACCAACACATTTTC 481
    |||||
DB 89 -----ValMetValLeuGlnLeuLysThrAspGlnAspProIleAsnPro 104
OY 482 TTCACACGAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 538
    |||||
DB 105 PheSerGlnValPheIleLeu-----ArgProAsnAsnGlnGlySerYrPhe 120
OY 539 ATTCGACGATGATGCTTCCGT 559
    |||||
DB 121 IleGlyAsnGlnIlePheArg 127

```

# RESULT 8

hypothetical protein K08F4.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T23479

R:Hendry, C.

submitted to the EMBL Data Library, January 1996

A:Reference number: Z19746

A:Accession: T23479

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-537 <WII>

A:Cross-references: EMBL:Z68879; PIDN:CAA93082.1; GSPDB:GN00022; CESP:K08F4.2

A:Experimental source: clone K08F4

C:Genetics:

A:Gene: CESP:K08F4.2

A:Map position: 4

A:introns: 66/3; 99/1; 140/3; 161/3; 338/1; 419/1

## Alignment Scores:

Pred. No.:	0.185	Length:	537
Score:	94.00	Matches:	32
Percent Similarity:	46.88%	Conservative:	28
Best Local Similarity:	25.00%	Mismatches:	48
Query Match:	4.67%	Indels:	20
DB:	2	Gaps:	6

US-09-763-902B-15 (1-1146) x T23479 (1-537)

```

OY 203 GCTGCTGAGAG-----TTTGTCAATATTTACTATGACACATGATGAATAAGA 250
    |||||
DB 96 AlaAlaGlnGlnValGlyGlyAlaPheCysHisGlnPheYrYrIleThrValSerGlnAsn 115

```



Db 19 phevalglulysyltyrhisnleuetyrlysserproserglnvalhsglnphery 38  
 QY 275 CTGGACAGAGCCACCTTAATATGGAATGGAAT-----CCTGTTTCAGCG 319  
 Db 39 leuaspaspservalleuenglarginproglyseraspolyglumetvalservallysser 58  
 QY 320 CTGGATGCCCTTAATAATTTTTTTGACACATTCCTTCAGTACAGTTCGACGTAATAG 379  
 Db 59 leuylalaileasn-----gluglnilemetserpheasptryglulileserlys 75  
 QY 380 TTGATTTGCCAACCACTTCATGACGACGACATCCCAACTACCTTCCTGTTG 439  
 Db 76 lileglnlleuethralaaspserglnalasertrymetasnolyvalylalthrleuval 95  
 QY 440 ACCAGTGAAGTCTGTAAGTTTGATGGAACAACAACATTTCTTCAACGACGACTCTCG 499  
 Db 96 thrglyleuethrvallysglarginmetarg---pheserqlnserphephe 114  
 QY 500 CTGACTGCTCAGTCCCTCCCAACATCTGTGTGAGACATTCGACAGTATGCTTCGCT 559  
 Db 115 leu-----valproleuasnnglysertyrphevalleuasnaspvalphearg 130  
 QY 560 TTT 562  
 Db 131 tyr 131  
 RESULT 11  
 S24382  
 nitrous-oxide reductase (EC 1.7.99.6) - Alcaligenes eutrophus  
 C:Species: Alcaligenes eutrophus  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S24382  
 R:Zumft, W.G.; Dreusch, A.; Loechelt, S.; Cuypers, H.; Friedrich, B.; Schneider, B.  
 Eur. J. Biochem. 208, 31-40, 1992  
 A:Title: Derived amino acid sequences of the nosZ gene (respiratory N(2)O reductase) from  
 A:Implications for the Cu(A) site of N(2)O reductase and cytochrome-c oxidase.  
 A:Reference number: S24382; M0ID:92380183; PMID:1324835  
 A:Accession type: DNA  
 A:Molecule type: DNA  
 A:Residues: 1-643 <ZUM>  
 A:Cross-references: EMBL:X65278; NID:938781; PIDN:CAA46383.1; PID:938782  
 C:Superfamily: nitrous-oxide reductase  
 C:Keywords: copper; metalloprotein; oxidoreductase  
 C:Keywords: copper; metalloprotein; oxidoreductase  
 F:589,624,628/Binding site: copper 1 (His, Cys, Cys) #status predicted  
 F:624,626,632/Binding site: copper 2 (Cys, His, Cys, Cys, His) #status predicted  
 Alignment Scores:  
 Pled. No.: 1,1 Length: 643  
 Score: 87.00 Matches: 40  
 Percent Similarity: 41.61% Conservative: 17  
 Best Local Similarity: 29.20% Mismatches: 62  
 Query Match: 4.32% Indels: 18  
 Db: 1 Gaps: 6  
 US-09-763-902B-15 (1-1146) x S24382 (1-643)  
 QY 137 ACTGTACAGCTCCAGATGGCAGCTCTGTGATTTT-----AAACTTAT 184  
 Db 361 ThrAlaThrValIleGlnLeuSerArgValIleGlnTyrPheaspolylysnglnulys 380  
 QY 185 GTAGATCAGCATGTAGACCTCTGAGAGTTT---GTCAATATTACTATGACCAATG 241  
 Db 381 leuaspaspalailevalnleuvalgluleuglyproleuhsnhrleuaphe 400  
 QY 242 GATTAAGAAGAGCGGACATAACGCTGTATCTGGACAGGCCACTTAATATGAAAT 301  
 Db 401 AspGlyArgGlyAsnAlaTyrThrThrLeuPheLeuaspserglnleuvallystrpahn 420  
 QY 302 GGAATATGCTTTTTCAGGCTGATGCCCTTAATATTTTTCAGACATTCCTCTACT 361  
 Db 421 -----leuaspalailelelysphehislysglyasplysasnala 434  
 QY 362 GAGTTCAGGTCATATATGTAGATTGC-----CAACGAGTTCATGACGCAACTCAG 415

Db 435 lystyrvalvalaspargleuaspserleuglntryglnproglyhisvalasnalaasergln 454  
 QY 416 TCCCAACACACAGTTCTTTGTGGACACAGTGAACGTGAAGTTTATGAAACACAA 475  
 Db 455 SerGlnThr-----valalaalaaspolylystyrleuvalalglycyslys--- 470  
 QY 476 CATTTCTTCAACGACGACTTCCTGCTGACTGCTCAGTCCACATCCCAACAT 526  
 Db 471 -----pheserlyaspargpheleuProvalglyProleuHisProcluan 486  
 RESULT 12  
 T31082  
 endo-1,4-beta-xylanase (EC 3.2.1.8) - Caldiceilus rufus sp.  
 C:Species: Caldiceilus rufus sp.  
 C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
 C:Accession: T31082  
 R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.  
 submitted to the EMBL Data Library, December 1997  
 A:Description: Family 10 and 11 xylanase genes from Caldiceilus rufus sp. Rte9B.1.  
 A:Reference number: Z20972  
 A:Accession: T31082  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1595 <MOR>  
 A:Cross-references: EMBL:AF036923; NID:92760904; PID:92760908; PIDN:AAB95325.1  
 C:Genetics:  
 A:Gene: xynB  
 C:Keywords: glycosidase; hydrolase  
 Alignment Scores:  
 Pled. No.: 1,66 Length: 1595  
 Score: 86.00 Matches: 32  
 Percent Similarity: 38.28% Conservative: 17  
 Best Local Similarity: 25.00% Mismatches: 50  
 Query Match: 4.27% Indels: 29  
 Db: 2 Gaps: 4  
 US-09-763-902B-15 (1-1146) x T31082 (1-1595)  
 QY 256 GGCACCTAACAGGCTGTATCTGACAGGCGACCTTAATATGAAATGCTGTTTC 315  
 Db 730 GlyLeuIleAspGlyValGlyLeuGlnGlnHisIleAsnValasp-SerProAlaVally 749  
 QY 316 AGGCGTGAATGCCCTTAATAATTTTTTTGACACATTCGCTTCAGTACAGTCCAGTCAA 375  
 Db 749 sculileglnuaspthrileasnleuapheSerThrIleProglyLeuGlnIleGlnIleth 769  
 QY 376 TATGTAGATTCGCAACACGATTCATGACGACGACGACGACGACGACGACGACGAC 435  
 Db 769 rgluleuasp-----ileserya 775  
 QY 436 TGTGACCACTGGAAGTGTGAAGTTGATGGAACAACAACATTTTC----- 481  
 Db 775 ltyrthrserSerThrGlnGlnIlytraspheleuProGlnAspIleMetileysglnal 795  
 QY 482 -----TTCACGCAACATTCCTCTGCTGACGCTGACGCTGACGCTGACGCTGACGCT 525  
 Db 795 aleuylsphePlysglnleuPheGlnmetleuylasglnHisSerAspArgIleThrAsnva 815  
 QY 526 TACTGTGTGAGATGATGACGATGATGCTTCGCTTTCAGATGCTGCTAGTAAAG 585  
 Db 815 lthrleutrpolylyleuylaspasp-----tyrprothrleuSerlyasparg 831  
 QY 586 GGGCAAAAGCTCATTCCTCATTT 607  
 Db 831 gserasntrpProleuLeuPhe 838  
 RESULT 13  
 VGTH79  
 E2 glycoprotein precursor - feline infectious peritonitis virus (strain 79-1146)  
 N:Alternate names: peplomer glycoprotein; spike glycoprotein  
 C:Species: feline infectious peritonitis virus



[illegible]

257 -----GCAC TAACCA GCGTGT ATCTG GACAAG GCGCACC TTATAT GG 298

Db 367 LeuHisValThrGlnGlyLeuValThrArgAlaTyrThrAspGlu-----LeuTyr 383

```

OY 299 AATGGAATGCTGTTTCAGGGCTGTGATGCC----- 328
    |||
    |||
    |||
Db 384 Asn---MetAlaLeuSerLysIleIleAlaValLeuArgAlaHisSerSerTyrCysThr 402
    |||
    |||
    |||
OY 329 -----CTAATATAT-----TTTGTGACACATTCCT 355
    |||
    |||
    |||
Db 403 AspProAspLeuValLeuGluLeuLysAsnLeuIleValIlePheAlaAspThrLeuGlu 422
    |||
    |||
    |||
OY 356 TCTAGTAGTTCAGGTCATATGTTAGTTCACACAGCTGATGACAGCAACTCAG 415
    |||
    |||
    |||
Db 423 GlyTyrGluPheSerValAsnArgLeu---PheAspLeuLeuPheGluIleArgAspGlu 441
    |||
    |||
    |||
OY 416 TCCCAACTACAGTCTGTTGTGACAGTGAAGTTCGAACTTGTATGCAACAACAA 475
    |||
    |||
    |||
Db 442 TyrAsnIuThrLeuLeuLysTyrAlaGlyIlePhe-----ArgAsp 456
    |||
    |||
    |||
OY 476 CATTCTTCAACGACGAAGCTTCCTGCTGACGCTCAGTCACACTCCC-----AAC 523
    |||
    |||
    |||
Db 457 IlePheGluGluAspAsnTyr-----SerProIleProIleGlySer 470
    |||
    |||
    |||
OY 524 AATCTCTGTGAGAGATTCGACAGTATTCCTGCTTTCAGAT-----TGCTCT 574
    |||
    |||
    |||
Db 471 GluGluGluIuTyrLysMetValIleSerLysPheProGlnAspProAspLeuGluLys 490
    |||
    |||
    |||
OY 575 AGTAGTTAAAGGGCAAAAGTCCATTCCTCATTTGTCATTCAGCATTCGAATGAAT 634
    |||
    |||
    |||
Db 491 GlnSerPheProLysLysPheProMetSer-----GlnSerValProLeuIleTyrIle 508
    |||
    |||
    |||
OY 635 TATGTGAATTTATTTGTTGATTGTAGACGACATATATATGCTGGAACATAATTT 688
    |||
    |||
    |||
Db 509 GlnValLysGluPhe-----IleTyrAlaSerLeuLysPheSerGlu 522
    |||
    |||
    |||
OY 689 -----CTTAATATTTTCTATTTCTCCGTAGACACCTTTCTA 724
    |||
    |||
    |||
Db 523 SerLeuHisArgSerSerThrGluIleAspSerPheLeuArgLysSerThrAsnLeuLeu 542
    |||
    |||
    |||
OY 725 GCACGCTCCAGTTGGAGCATTTGCCCTCTAAGACCTTAAACATATTTTTCATGCT 784
    |||
    |||
    |||
Db 543 LeuThrArgIleLeuSerSerCysLeuLeuAsnLeuLeuArg-----LysPro 558
    |||
    |||
    |||
OY 785 TATATACATTCCTATATGACATTCCTATATATATATTAACACATGATCTTGTAATA 844
    |||
    |||
    |||
Db 559 HisIleGlyLeuThr---GluLeuValGlnIleIleIleAsnThr----- 572
    |||
    |||
    |||
OY 845 CATACCTCAGCTGAA 859
    |||
    |||
    |||
Db 573 ---ThrHisLeuGlu 576
    |||
    |||
    |||

```

## RESULT 15

E96716

probable RNA-binding protein F23010.17 [Imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: E96716

R:Proteologs: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hutzar, L.

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: E96716

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-427 &lt;STO&gt;

A:Cross-references: GB:AE005173; NID:g7705097; PIDN:AAF67776.1; GSPDB:GN00141

C:Genetics:

A:Gene: F23010.17

A:Map position: 1

```

Alignment Scores:
Pred. No.: 2.45
Score: 83.50
Percent Similarity: 41.40%
Best Local Similarity: 23.57%
Query Match: 4.15%
DB: 2
Gaps: 6

US-09-763-902b-15 (1-1146) x E96716 (1-427)

OY 185 GTAGATCAGCAGCATTAAGCTGCTGAGAGAGTTGTCATATATTACTATGACACATGAT 244
    |||
    |||
    |||
Db 7 ValProSerAlaGlnAspIleAlaGluPheValArgGlnTyrHisValLeuGly 26
    |||
    |||
    |||
OY 245 AAAGAAGAGGGGACATACACAGGCTGTATCTGACAGCAAGCCACCTTAATA----- 295
    |||
    |||
    |||
Db 27 GlnLeuProHisGluAlaArgArgLeuTyrValAspAlaSerValValSerArgProAsp 46
    |||
    |||
    |||
OY 296 TGGATGGAATGCTGTTCA-----GGCTGATGCCCTTAATATTTTTCACACA 349
    |||
    |||
    |||
Db 47 ValThrGlyThrMetMetSerPheThrSerValGluAlaIleAsnLysHisIleLeuSer 66
    |||
    |||
    |||
OY 350 -----TTGCCCTTCTAGAGTTCAGAGTTCATATGTTAGTTCACACAGTTCATGAG 403
    |||
    |||
    |||
Db 67 CysAspPheGluAsnThrLysPheGluValLeuSerValAspSerGlnAsnSerLeuGlu 86
    |||
    |||
    |||
OY 404 CAAGCAACAGTACGCCAAGCTCTGTTGTGTGACCACTGAGACTGGAAGTTGAT 463
    |||
    |||
    |||
Db 87 AspGly-----IlePheIleMetValIleGlyPheMetThrGlyLys 100
    |||
    |||
    |||
OY 464 GGAACAAACACATTTCTTCACACACACTCTGCTGACTGCTGACTGCTGACTGCTGACT 523
    |||
    |||
    |||
Db 101 AspAsnGlnArgArgLysPheSerGlnMetPheThrLeuAlaArgGlnAsnThr----- 118
    |||
    |||
    |||
OY 524 AATACCTGTGAGAGATTCGACAGTATTCCTCCGTTTCAAGATTGCTAGAGTTAA 583
    |||
    |||
    |||
Db 119 -----LeuValValLeuAsnAspMetLeuArgTyrValAspGlnGluAspSerSer 135
    |||
    |||
    |||
OY 584 AGGGCAAAAGTCCATTCATTTGTCATTTAGTTCACGACATTAAT 634
    |||
    |||
    |||
Db 136 ThrThrGluThrProCysGlu-----ProValThrGluIle 147
    |||
    |||
    |||

```

Search completed: June 24, 2003, 19:39:23

Job time : 48.4874 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_n2p model

Run on: June 24, 2003, 19:31:14 Search time 43.497 Seconds  
(without alignments)  
5701.778 Million cell updates/sec

Title: US-09-763-902B-15

Perfect score: 2013  
Sequence: 1 taccctcgagagagatgg.....tttaagtaattttgtttt 1146

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Xgapop 10.0	Xgapext 0.5
Xgapop 6.0	Xgapext 7.0
Delop 6.0	Delext 7.0

Searched: 417779 segs, 108206813 residues

Total number of hits satisfying chosen parameters: 835558

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlp  
-MODE-frame-n2p.model -DEV-xlp  
-Q/cgn2\_1/USPTO/US09763902/runat\_24062003\_130343\_12522/app\_query.fasta\_1\_2234  
-DB-published\_Applications\_AA -QPM-fastlan -SUFFIX-n2p.rapb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62  
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR-score-pct -THR-MAX=100  
-THR-MIN=0 -ALIGN=15 -MODE-LOCAL -OUTFMT=ptc -NOR-ext -HEARSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USPR-US09763902.eccn\_1\_1\_24\_etunal\_24062003\_130343\_12522  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPELCK=100  
-LONGLOG -DEV-TIMEOUT=120 -WARN-TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications, AA:

1:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2:	/cgn2_6/ptodata/2/pubpaa/PCIT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/PCITUS_PUBCOMB.pep.*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	727	36.1	162	10	US-09-764-864-1180
2	584.5	29.0	164	10	US-09-764-864-1181
3	91.5	4.5	170	9	US-09-791-932-107
4	88.5	4.4	449	9	US-10-074-475-272

	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
	87	85	83.5	83.5	82	81.5	81	81	81	79.5	78.5	78.5	78.5	78.5	78	78	78	76.5	76	75	75	75	75	74.5	74	74	73.5	73	73	72.5	71.5	71.5	71	71	71	71	70.5	70.5	70.5	70.5	70	
	4.3	4.2	4.1	4.1	4.1	4.0	4.0	4.0	4.0	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.8	3.8	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.5	3.5	3.5	3.5	3.5	3.5	
	1454	748	1127	1158	752	9	358	358	358	358	1454	1454	1454	1454	1454	1454	1454	1369	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	1354	
	US-09-854-799-22	US-09-854-799-24	US-09-815-242-5373	US-09-815-242-5373	US-10-176-847-52	US-10-077-130-5	US-09-852-209A-13	US-10-131-600-13	US-10-139-876-2	US-09-854-799-32	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11	US-10-156-239-11

#### ALIGNMENTS

RESULT 1  
US-09-764-864-1180  
Sequence 1180, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764,864  
CURRENT FILING DATE: 2001-01-17  
PRIOR application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1180  
LENGTH: 162  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-764-864-1180

#### Alignment Scores:

Pred. No.:	3	516-80	Length:	162
Score:	727.00		Matches:	137
Percent Similarity:	100.00%		Conservative:	0
Best Local Similarity:	100.00%		Mismatches:	0
Query Match:	36.12%		Indels:	0
DB:	10		Gaps:	0

US-09-763-902B-15 (1-1146) x US-09-764-864-1180 (1-162)



PRIOR FILING DATE: 2001-02-13  
NUMBER OF SEQ ID NOS: 295  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 272  
LENGTH: 449  
TYPE: PRT  
ORGANISM: Homo sapien  
US-10-074-475-272

Alignment Scores:  
Pred. No.: 0.161 Length: 449  
Score: 88.50 Matches: 34  
Percent Similarity: 42.408 Conservative: 19  
Best Local Similarity: 27.208 Mismatches: 63  
Query Match: 4.408 Indels: 9  
DB: 9 Gaps: 4

US-09-763-902B-15 (1-1146) x US-10-074-475-272 (1-449)

QY 212 GAGTTTGCAATATTACTATGACACAAATGATAAAGAGACGGGCACTAACCGGCTG 271  
Db 14 GlnPheValAlaArgIleTyrThrLeuLeuAsnLysAlaProGluTyrLeuHisArgPhe 33

QY 272 TACTGGAACAAGCCCACTTAATATGAGATGGAATGCTGTTCAGG----- 319  
Db 34 TyrGlyArgAsnSerSerTyrValHisGlyValAlaSerArgLysProGlnGlu 53

QY 320 ---CTGAGAGCCCTAAATATTTTGTGACACATTCGCTCTAGTGAATTCACAGTCAT 376  
Db 54 AlaValTyrGlyGlnAsnAspIleHisLysValLeuSerLeuAsnPheSerGlyCys 73

QY 377 ATGTAGATTGCCAACCACTTCATGAGCAAGCACTGATCCCAACTACAGTCTTGT 436  
Db 74 HisThrLysIleLeuArgHisValAspAlaHisAlaThrLeuSerAsp---GlyValVal 92

QY 437 GTACCAAGTCGACAGCTGTGAAGTTGATGGAACAACAACACTTTCCTTCAACCAAGTTC 486  
Db 93 GlnValMetGlyLeuLeuSerSerAsnSerGlyGlnProGluArgLysPheMetGlnThrPhe 112

QY 497 CTGCTAGACGTGCTAG---TCCACGCTCCCAACAATACTGTGTGGAAGATTGCAAGTATGC 553  
Db 113 ValLeuAlaProGluLysSerValProAsnLys-----PheTyrValHisAsnAspMet 130

QY 554 TTCGGTTTCAAGAT 568  
Db 131 PheArgTyrGluAsp 135

RESULT 5  
US-09-854-799-22  
Sequence 22, Application US/09854799  
Patent No. US20020115064A1  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King Of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/854,799  
FILING DATE: 14-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/392,459  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5090  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ. ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 22:  
US-09-854-799-22

Alignment Scores:  
Pred. No.: 0.402 Length: 1454  
Score: 87.00 Matches: 69  
Percent Similarity: 33.77% Conservative: 34  
Best Local Similarity: 22.62% Mismatches: 112  
Query Match: 4.32% Indels: 90  
DB: 10 Gaps: 17

US-09-763-902b-15 (1-1146) x US-09-854-799-22 (1-1454)

OY 260 CTACACGAGCTGATCTGACACAGCCACTTAATGAAATGCTGTTTCAGG 319  
Db 211 ValThleuLeuYrserArGserThAlaThrTpoluYrserAlaAlaTyrAla 230  
OY 320 CTGGATGCCCTAAATAATTTTTCACACATTCCTTCAGAGTTCAGTCAATANG 379  
Db 231 TyrGlnGlyValSerAsnPe-----ThrTyrTyrTyrLeuAsnAsn 244  
OY 380 TTAGAT-----TGCCACAGCTTCATGACAGACAGCACTGCTCC 418  
Db 245 ThrAsnGlyLeuYsthrTyrGluLeuGlySgluAsp---TyrGluHisCysThrGlyTyr 263  
OY 419 CAATACACAGCTCTGTTGTGACAGTGAAGTGTGATGATGGAACAAACAACAT 478  
Db 264 AlaThrAsnValPheAlaProThrSerGlyGlyTyrIleProAspGly-----Phe 280  
OY 479 TTCTTCACACAGACTCTGCTGACTGCTGCTCAGT----- 517  
Db 281 SerPheAsnAsnTrpPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
OY 518 CCACAC-----AATACGTGTGGAAGTTGCAGT----- 547  
Db 301 ThrAsnGlnProLeuLeuIleAsnCysLeuTrpProValProSerPheGlyValAlaAla 320  
OY 548 -----GATGCTTCGCTTTTCAAGATGCTAGTAAAGGGCAAAAGTCCATTC 601  
Db 321 GlnGluPheCysPheGluGlyAlaGlnPheSerGlnCys-----AsnGlyVal 336  
OY 602 TCATTGGTCCATTAGTCCAGCAATGAATTAATTAATGATATAT----- 646  
Db 337 SerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnPheThrAlaAspValGln 356  
OY 647 -----TTT 649  
Db 357 SerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyGlyValIleLeuGluIle 376  
OY 650 GATTGTAGACACATAATATATGCTGACAACTAATTTCTTAATATTTCTATCCG 709  
Db 377 SerCysTyrSerAspThrValSerGluSerSer-----TyrSerTyrGly 392  
OY 710 TCAGCACCTTTTCTAGCAGCTGCCAGTTGGACCAATGCCCCCTTA-----AGA 757  
Db 393 GluIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeuTyrAsnGlyThr 412

OY 758 GCTTAACATATTTTATACATGCTTAATATACATTCACATGACATCTTATATA 817  
Db 413 AlaLeuLysTyrLeuGlyThrLeuPro-----ProSerValLysIleAlaIle----- 429  
OY 818 AATTAACACATGATCTTGATGATACATACATCTGATGACCCAGCTTATGCAAAA 877  
Db 430 -----SerLysTrpGlyHisPheTyr-----IleAsnGly 439  
OY 878 TAAATCTTTTATATATATATATATATGATGATGATGATGATGATGATGATGAT 934  
Db 440 TyrAsnPhePheSerThrPheProIleGlyCys--IleSerPheAsnLeuThrThrGlyVal 459  
OY 935 GAATGAGATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 994  
Db 459 IAserGlyAlaPheTrpThrIleAla---TyrThrSerTyrThrGluAlaLeuValGlnVal 478  
OY 995 TT---AACACT 1002  
Db 478 AlaGluAsnThr 481

RESULT 6  
US-09-854-799-24  
Sequence 24, Application US/09854799  
Patent No. US20020115064A1  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESS: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/854,799  
FILING DATE: 14-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/392,459  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5090  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-09-854-799-24

Alignment Scores:  
Pred. No.: 0.537 Length: 748  
Score: 85.00 Matches: 69  
Percent Similarity: 33.77% Conservative: 34  
Best Local Similarity: 22.62% Mismatches: 112  
Query Match: 4.22% Indels: 90

[illegible]

APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELTRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 5373  
LENGTH: 1127  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5373

Alignment Scores:				
Pred. No.:	0.973	Length:	1127	
Score:	83.50	Matches:	30	
Percent Similarity:	39.34%	Conservative:	18	
Best Local Similarity:	24.59%	Mismatches:	45	
Query Match:	4.15%	Indels:	29	
	10	Gaps:	6	

US-09-763-902B-15 (1-1146)	x	US-09-815-242-5373 (1-1127)
QY 170 GATTTTAAACCTTATGTAGATCAGCGCATGAGCTGCTGAGACTTTGTCAATATTTTAC		2295
123 Asphrheylsystrylaa--ValThrProGlnHisLeuGlnSerPheIe-----		1380
QY 230 TATGACACAATGGATTAAGAAAGACGGCAGTACACAGGCTGTATCTGGACAAAGCCACC		2875
139 -----AlasprlysasmetGlnThrAgtThrIAsnLysLeuGlnAspIleLa		1555
QY 290 TTAATATGCG-----ATGGAATAGCTGTTTACGGGCTGGATGCC		3225
156 LeuIlelytRgtGclurPheGlnLarGllLeglInasngLurPheIleThrGlyGluAspLa		1757
QY 329 CTAATAATTAATTTTGTGACACATTTGCGCTCTAGTAGAGTCCAGGTCATATATTTTGAATGTC		3885
176 LeuGlnItyrPheIleAspCysMetProLysSerGluTyr-----LeuLysCys		1917
QY 389 CAACCAAGTTCATGAGCAAGCAACATCACTGCCCAACTACAGTT--CTTGTGTGACCAAGT		4455
Db 192 AlasprIleleYrIleAspGlyPheNHisAsnPheserThrIleGluTyrLeuIleIleLys		2111
QY 446 GGAAGCTGTGAAGTTT-----GATGCAACAACAA		4755
Db 212 GlyLeuIlelytRgtValLysSerValThrIleIleLeuThrIleAspGlyAsnHisAsp		2311
QY 476 CATTTG 481		
Db 232 GlnPhe 233		







DB 193 glycylserylserineprothylprothylprothylserinelelle-arganysertl 212  
QY 222 TTGACAACTCCTACAGAGCTTACATGCTGATCTACATGTTTAAATCCAGAGAC 163  
DB 212 e---GlnthPrrogluInspRlucysPrroHlsertLysleucysProileaspme 231  
QY 162 GTGGCCATCGGAGACCTTGAGCAGTGTCTGCGAGTCCGGCTACTACT----- 108  
DB 231 tleutPrasPasnThLysCysValleuGlnaspGluThProleuProglYth 251  
QY 107 -----CCTCAC----- 102  
DB 251 rgluaspHlsertTyrlleuGlnIncluprothleucysGlyPrroHlsmethrheasp6l 271  
QY 101 -----GTGTCGGTCCGCA 88  
DB 271 uasPargCysglucysValCyslysalaProCysProglYaspLeuileGlnHlsProgl 291  
QY 87 GGGCTGGGAAATGATCAATCTATGAGAGAGAGAAAC---CGCCAAACGCAAAAT 31  
DB 291 uasCysSerCyspheglucyslysGluSerLeuGluSerCysCysGlnLysHlsYstl 311  
QY 30 TTTCACCCCTCCCATTCCTCTCC 9  
DB 311 ephenHlsProaspThrcysSer 318  
RESULT 12  
US-10-131-600-13  
Sequence 13, Application US/10131600  
Publication No. US20030082670A1  
GENERAL INFORMATION:  
APPLICANT: ERIKSSON, Ulf  
APPLICANT: ERIKSSON, Karin  
APPLICANT: LEE, Xuri  
APPLICANT: PONTEN, Annica  
APPLICANT: UUTELA, Marko  
APPLICANT: ALITALO, Karl  
APPLICANT: OESTMAN, Arne  
APPLICANT: HELDIN, Carl-Henrik  
APPLICANT: BETSHOLTZ, Christer  
TITLE OF INVENTION: PLATELET-DERIVED GROWTH FACTOR C, DNA CODING  
TITLE OF INVENTION: THEREFOR, AND USES THEREOF  
FILE REFERENCE: 09-410349-Eriksson et al-1064-44740  
CURRENT FILING DATE: 2002-04-25  
PRIOR APPLICATION NUMBER: US/10/131,600  
PRIOR FILING DATE: 1999-09-30  
PRIOR APPLICATION NUMBER: 60/108,109  
PRIOR FILING DATE: 1998-11-12  
PRIOR APPLICATION NUMBER: 60/110,749  
PRIOR FILING DATE: 1998-12-03  
PRIOR APPLICATION NUMBER: 60/113,002  
PRIOR FILING DATE: 1998-12-18  
PRIOR APPLICATION NUMBER: 60/135,426  
PRIOR FILING DATE: 1999-05-21  
PRIOR APPLICATION NUMBER: 60/144,022  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 13  
LENGTH: 358  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-131-600-13  
Alignment Scores:  
Pred. No.: 1.22 Length: 358  
Score: 81.00 Matches: 38  
Percent Similarity: 34.52% Conservative: 20  
Best Local Similarity: 22.62% Mismatches: 67  
Query Match: 4.00% Indels: 43  
DB: 9 Gaps: 5

US-09-763-902b-15 (1-1146) x US-10-131-600-13 (1-358)  
QY 387 CAATCTACATATGTAAGCTGGAAGCTACAGAGCAATGCTCAAAATATTTAGS 328  
DB 153 GlnGlnGlnValMetCysMetAsnThrsertHrsertTyrlleuSerGlnleuPheglu 172  
QY 327 GCATCCAGCCCTTAACAGAGATTTCCATTCATTAAGTGGCCCTTGCACAGC 268  
DB 173 lleserValProleuThrsertValProgluLeuValProValLysllealaAsnHlsThr 192  
QY 267 -----CTGTTAGTCCGCCCTCTTTTATCCATGCTCATAGTAATA 223  
DB 193 glyCylserylserineprothylprothylprothylserinelelle-arganysertl 212  
QY 222 TTGACAACTCCTACAGAGCTTACATGCTGATCTACATGTTTAAATCCAGAGAC 163  
DB 212 e---GlnthPrrogluInspRlucysPrroHlsertLysleucysProileaspme 231  
QY 162 GTGGCCATCGGAGACCTTGAGCAGTGTCTGCGAGTCCGGCTACTACT----- 108  
DB 231 tleutPrasPasnThLysCysValleuGlnaspGluThProleuProglYth 251  
QY 107 -----CCTCAC----- 102  
DB 251 rgluaspHlsertTyrlleuGlnIncluprothleucysGlyPrroHlsmethrheasp6l 271  
QY 101 -----GTGTCGGTCCGCA 88  
DB 271 uasPargCysglucysValCyslysalaProCysProglYaspLeuileGlnHlsProgl 291  
QY 87 GGGCTGGGAAATGATCAATCTATGAGAGAGAGAAAC---CGCCAAACGCAAAAT 31  
DB 291 uasCysSerCyspheglucyslysGluSerLeuGluSerCysCysGlnLysHlsYstl 311  
QY 30 TTTCACCCCTCCCATTCCTCTCC 9  
DB 311 ephenHlsProaspThrcysSer 318  
RESULT 13  
US-10-139-876-2  
Sequence 2, Application US/10139876  
Patent No. US20020123481A1  
GENERAL INFORMATION:  
APPLICANT: Oliviero, Salvatore  
TITLE OF INVENTION: C-Fos induced Growth Factor (Figf) And Dna Encoding Same  
FILE REFERENCE: 35784/205172  
CURRENT FILING DATE: 2002-05-07  
PRIOR APPLICATION NUMBER: US/10/139,876  
PRIOR FILING DATE: 1998-03-18  
PRIOR APPLICATION NUMBER: 09/043,476  
PRIOR FILING DATE: 1998-03-18  
PRIOR APPLICATION NUMBER: PCT/IB96/0113  
PRIOR FILING DATE: 1996-09-30  
PRIOR APPLICATION NUMBER: GB9612368.2  
PRIOR FILING DATE: 1996-06-13  
PRIOR APPLICATION NUMBER: GB9519928.7  
PRIOR FILING DATE: 1995-09-29  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 358  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-139-876-2  
Alignment Scores:  
Pred. No.: 1.22 Length: 358  
Score: 81.00 Matches: 38  
Percent Similarity: 34.52% Conservative: 20  
Best Local Similarity: 22.62% Mismatches: 67  
Query Match: 4.00% Indels: 43  
DB: 12 Gaps: 5

US-09-763-902B-15 (1-1146) x US-10-139-876-2 (1-358)

OY 387 CAATCTACATATATGACCTCGGAACCTACTAGAACCAATGTGTCAAAAATATATTAGG 328B  
Db 153 GILGLUGLlyAlMetetysmetasnthrSerthrsertYrIleSerlysgInleuphglu 172B  
OY 327 GCATCCAGCCCGTAAACACAGCATTTCCATTCATTCATTAAGAGGCGCCTTCCAGATTACAGC 266B  
Db 173 lIleSerValProleuthrSerValProgluleuValProValLySIleAlaasnHsthr 192B  
OY 267 -----CTGGTAGTGCCGCTCTTTATCCATTCGTCTCATTAATAATA 223B  
Db 193 GLyCyLyScysleuProthrglyProhtrghnSProtyrserIlelle-ArgArgserI 212B  
OY 222 TTGACAAATCCTCTACAGCAGCTCTAATGCGTGATCTACATAGATTAAATTCAGAGAC 163B  
Db 212 e--GlnthrProglIngluAspGLyCysProHisSerlyslsYleuCsProIleasPme 231B  
OY 162 GTGGCCATCTGGACCTTTAGCAGTGTGTGGCAATGTCGGCGTCATCACT----- 108B  
Db 231 tleuTPaspasnthrLyScysLyScysValleuInaspgluThrProleuProglYrth 251B  
OY 107 -----CCTCAC----- 102B  
Db 251 rgluAspHisSerTyrleuGlnGluProThrleuCYselyProHisMetThrPheaspGl 271B  
OY 101 -----GTGTCCGGTCCGCA 88B  
Db 271 uAspArgCySGlucyValCyLyAlaAlaProCySProglYaspIleuIleGlnHSProgl 291B  
OY 87 GGGCTCGCGAATGATCAATACTTATTAGAGAGCGCAACC---CGCCAAACGCACAAAT 31B  
Db 291 uAsnCysSerCySPheGluCyLySGlSerleuGlnSerleuGlnSerCySGlnYshISlySI 311B  
OY 30 TTTCCACCCCTCATCTCTCC 9B  
Db 311 ePheHISProaspThrCySSer 318B

RESULT 14  
US-09-854-799-32  
Sequence 32, Application US/09854799  
Patent No. US20020115064A1  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/854,799  
FILING DATE: 14-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/392,459  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION

	TELEPHONE: (215) 270-5015
	TELEFAX: (215) 270-5090
	INFORMATION FOR SEQ ID NO: 32:
	SEQUENCE CHARACTERISTICS:
	LENGTH: 1454 amino acids
	TYPE: amino acid
	TOPOLOGY: linear
	MOLECULE TYPE: protein
	SEQUENCE DESCRIPTION: SEQ ID NO: 32:
	US-09-854-799-32
	Alignment Scores:
	Pred. No.:           3 36           length:           1454
	Score:               79 50           Matches:          77
	Percent Similarity:  34.15%       Conservative:     35
	Best Local Similarity: 23.48%      Mismatches:      121
	Query Match:           3 95%       Indels:           95
	DB:                    10           Gaps:            20
US-09-763-902B-15 (1-1146) x US-09-854-799-32 (1-1454)	
OY	209 GAGGAGATTGTCTCAATATTAC-----TANGACACAATGGATAAA 244
Dd	187 Aspolunphevalrhrlatrylleserlglyarisertryanstrpnslleasnasn 206
OY	248 AGAAGACGGCCACTAACCGAGCTGTAATTCGCACAAAGCCACCCTPATATGGAATGGAAT 307
Dd	207 TTPPheasnasnvalThrleuleurySerArgerSerthrAlathrtprglutySer 226
OY	308 GCAGTTTCACAGCGCTGGATGCCCAAATATTT-----TTTGAC 346
Dd	227 AlAlalatyValtyrlnglnlyalserasnphetrTyTyrysleuasnsnthasn 246
OY	347 ACATTCGCTTCAGTAGAGTTCCAGGTCAATATGTAATGATGCCAACAGTTACGACAA 406
Dd	247 GlyleuushrrthyrgluPhe-----Cysgluasp--Tyglutyr 259
OY	407 GCACCTACGCCCCAACCTACAGTCTCTGTGTGACCACTGGACAGTCGTGAAGTTGATGGA 466
Dd	260 CysThrcltyrAlathrasnvalPhealaProthvalGlgyLyTyrlleproaspoly 279
OY	467 AACAAACAACTTTCTTCCACCAAGACTTCCTGCTGACTGCACAGTCCACT----- 517
Dd	280 -----PheSerPheasnsntrPhleuleuthrasnSerSetThPhevalSer 296
OY	518 -----CCCAAC-----AATACTGTGTGGAGATTGCAAGT-- 547
Dd	297 GlyArphevalThrasnnglnProleuleuValasnCySleutrPrrovalProSerPhe 316
OY	548 -----GATTGCTTCGCTTTTCCAGATTGGTCTAGATTAAAGGGC 585
Dd	317 GlyValAlaIagIngluphecypnecluglYalaglInPheSerGlnCys----- 333
OY	590 AAAGTCGCACTTCATTTGGTGCATTAAGTCCAGCAATGGAATTTATGCAATAT-- 644
Dd	334 ---SerelyValserLeuasnsnthrValAspvalIlleArgPheasnlleuasPheThr 355
OY	647 -----TTTGATTGTAGAGCACT----- 664
Dd	353 AlaAspValcInserylmetyglYalAthrValPheSerleuasnthThrclyGlyVal 372
OY	665 -----ATAATATGT-----GCTGAACCTAAATTTCTTAAATATTT 700
Dd	373 IleleuGlualyserCystyranspThrValSerGIusSerPhe-----Tyr 369
OY	701 CTATTCCTGTCACACCTTTCTAGACAGTCGCACATTGGAGCATGAGCCCTCA----- 754
Dd	390 SerTyrclylulterophbeglylethrAspglyProArglyrcystryValleuty 409
OY	755 -----AGACCTTAAACTATTTTTTACATGCTTATATACATTCACCTAATGACATT 808
Dd	410 AsncglyThrAlaleuushrrleuenglYthrleupro-----ProSerVallysgluile 427



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:27:58 ; Search time 14.789 Seconds

(without alignments)  
4559.976 Million cell updates/sec

Title: US-09-763-902B-15

Perfect score: 2013  
Sequence: 1 taccctccggagagagatggg.....tttaagtaattttgtttt 1146

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus.n2p.model -DEV=xlp  
-Q/cgn2.1/USPTO\_SPOOL/US09763902/runat\_24062003.130342.12495/app\_query.fasta\_1.2254  
-DB=Issued Patents\_AA -QFMT=fastin -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -SMART=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09763902 -ECGN=1.1.42 -runat\_24062003.130342.12495 -NCPU=6 -ICPU=3  
-NO\_MMAP -LANG=EN -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Issued Patents\_AA:  
1: /cgn2.6/prodata/1/1aa/5A.COMB.pcp:\*  
2: /cgn2.6/prodata/1/1aa/5B.COMB.pcp:\*  
3: /cgn2.6/prodata/1/1aa/6A.COMB.pcp:\*  
4: /cgn2.6/prodata/1/1aa/6B.COMB.pcp:\*  
5: /cgn2.6/prodata/1/1aa/PCUTUS.COMB.pcp:\*  
6: /cgn2.6/prodata/1/1aa/BackKills1.pcp:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	4.3	1454	US-08-392-459-22	Sequence 22, Appl
2	87	4.3	1454	PCT-US91-08525-22	Sequence 22, Appl
3	87	4.3	1454	PCT-US93-04384-2	Sequence 2, Appl
4	87	4.3	1454	PCT-US93-04384-47	Sequence 2, Appl
5	86	4.3	544	US-08-591-685-9	Sequence 47, Appl
6	86	4.3	1454	PCT-US93-04384-44	Sequence 44, Appl
7	86	4.3	1454	PCT-US93-04384-45	Sequence 44, Appl
8	85	4.2	748	US-08-392-459-24	Sequence 24, Appl
9	85	4.2	748	PCT-US91-08525-24	Sequence 24, Appl
10	85	4.2	748	PCT-US93-04384-4	Sequence 4, Appl
11	81	4.0	358	US-08-915-795-8	Sequence 8, Appl
12	80.5	4.0	1454	PCT-US93-04384-43	Sequence 43, Appl

13	80.5	4.0	1454	PCT-US93-04384-46	Sequence 46, Appl
14	80	3.9	321	US-08-915-795-9	Sequence 9, Appl
15	79.5	3.9	1454	US-08-392-459-32	Sequence 32, Appl
16	79.5	3.9	1454	PCT-US91-08525-32	Sequence 32, Appl
17	79.5	3.9	1454	PCT-US93-04384-12	Sequence 12, Appl
18	79.5	3.9	1454	PCT-US93-04384-48	Sequence 48, Appl
19	79	3.9	1168	US-08-620-717A-9	Sequence 9, Appl
20	78	3.9	748	US-08-392-459-28	Sequence 28, Appl
21	78	3.9	748	US-08-392-459-34	Sequence 34, Appl
22	78	3.9	748	PCT-US91-08525-28	Sequence 28, Appl
23	78	3.9	748	PCT-US91-08525-34	Sequence 34, Appl
24	78	3.9	748	PCT-US93-04384-6	Sequence 6, Appl
25	78	3.9	1454	US-08-392-459-26	Sequence 26, Appl
26	78	3.9	1454	PCT-US91-08525-26	Sequence 26, Appl
27	78	3.9	1454	PCT-US93-04384-8	Sequence 8, Appl
28	78	3.9	1454	PCT-US93-04384-16	Sequence 16, Appl
29	78	3.8	2325	US-08-417-089-6	Sequence 6, Appl
30	78	3.8	2325	US-08-695-651-6	Sequence 6, Appl
31	78	3.8	2325	US-08-930-285-6	Sequence 6, Appl
32	78	3.8	2325	US-08-695-421-6	Sequence 6, Appl
33	78	3.8	2325	US-08-697-826A-10	Sequence 10, Appl
34	77.5	3.8	530	US-09-180-852-2	Sequence 2, Appl
35	77.5	3.8	1451	US-08-308-872B-4	Sequence 4, Appl
36	76.5	3.8	3169	US-09-453-702B-257	Sequence 257, App
37	76	3.8	238	US-08-858-207A-420	Sequence 420, App
38	75	3.7	325	US-08-915-795-3	Sequence 3, Appl
39	75	3.7	354	US-08-915-795-5	Sequence 5, Appl
40	75	3.7	590	US-08-756-317-12	Sequence 12, Appl
41	75	3.7	1335	US-09-134-001C-3716	Sequence 3716, Ap
42	74.5	3.7	401	US-08-492-459-28	Sequence 28, Appl
43	74.5	3.7	401	US-08-716-873-42	Sequence 42, Appl
44	74.5	3.7	401	US-09-368-431-42	Sequence 42, Appl
45	74.5	3.7	401	US-09-414-006-28	Sequence 28, Appl

#### ALIGNMENTS

RESULT 1  
US-08-392-459-22  
; Sequence 22, Application US/08392459  
; Patent No. 6280974  
; GENERAL INFORMATION:  
; APPLICANT: SmithKline Beecham, Corporation  
; TITLE OF INVENTION: Recombinant Peline Coronavirius S  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/392,459  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:



380 TAGAT-----TGCCACGAGTTCATGAGCAAGCACTGATCC 418  
245 ThrAsnGlyLeuLysThrTyrGluLeuGlySgluAsp---TyrGluHisCysThrGlyTyr 263  
419 CAACACTACAGTCTTGTGTGACAGCTGGAAGTGTGAGTGAACAAACACAT 478  
264 AlaThrAsnValPheAlaProThrSerGlyTyrIleProAspGly-----Phe 280  
479 TTCTTACACAGACACTTCTGCTGACTGCTGACTCCACT----- 517  
281 SerPheAsnAsnTrpPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
518 CCCAAC-----AATACTGCTGGAACATTCGACAT----- 547  
301 ThrAsnGlnProLeuLeuIleAsnCysLeuTrpProValProSerPheGlyValAlaAla 320  
548 -----GATTGCTCCGCTTTTCAAGATTGCTAGTAAAGGGCAAAAGTCATTC 601  
321 GlnGluPheCysPheGluGluGlyAlaGlnPheSerGlnCys-----AsnGlyVal 336  
602 TCATTGGTCCATTACTTCACAGCAATGAATTATGATATAT----- 646  
337 SerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnPheThrAlaAspValGln 356  
647 -----TTT 649  
357 SerGlyMetGlyValAlaThrValPheSerLeuAsnThrThrGlyGlyValIleLeuGluIle 376  
650 GATTGTAGACAGCACTTAATATGCTGGAACATAATTTCTTAATTTCTATTCCTG 709  
377 SerCysTyrSerAspThrValSerGluSerSer-----TyrSerTyrGly 392  
710 TCAGCACCTTTTTCACAGCTGCCAGCTTGGACATTCGCCCTCA-----AGA 757  
393 GlnIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeuTyrAsnGlyThr 412  
758 GCTTAAACTATTTTTCATGCGCTTATATACATTCACATGATCTTATATA 817  
413 AlaLeuLysTyrLeuGlyThrLeuPro-----ProSerValLysGluIleAlaIle--- 429  
818 ATATTAAACACATGATCTGTACTAATACATCTGATGACCCAGCTTATGCAAAA 877  
430 -----SerLysTrpGlyHisPheTyr-----IleAsnGly 439  
878 TAAATCTTTTATATATATATGATGAGTGCAGACATAATAC---ACTGTGGAA 934  
440 TyrAsnPhePheSerThrPheProIleGlyCys---IleSerPheAsnLeuThrThrGly 459  
935 GAAGTGAGAGTTTGTGTATAGCTTAATTTCTAGTAAACAGCTGCTGTTTCAG 994  
459 IAserGlyAlaPheTrpThrIleAla---TyrThrSerTyrThrGluAlaLeuValGln 478  
995 TT---AACACT 1002  
478 alGluAsnThr 481

RESULT 3  
PCT-US93-04384-2  
Sequence 2, Application PC/TUS9304384  
GENERAL INFORMATION:  
APPLICANT: Miller, Timothy J.  
APPLICANT: Kieffer, Sharon  
APPLICANT: Reed, Albert Paul  
APPLICANT: Jones, Elaine V.  
TITLE OF INVENTION: Compositions and Methods for Vaccination  
TITLE OF INVENTION: Against Coronaviruses  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation - Corporate  
ADDRESSEE: Patents  
STREET: 709 Swedeland Road  
CITY: King of Prussia

STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04384  
FILING DATE: 19930507  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,171  
FILING DATE: 08-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Schreck, Patricia A.  
REGISTRATION NUMBER: 33,777  
REFERENCE/DOCKET NUMBER: SBC H85009-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-04384-2

Alignment Scores:  
Pred. No.: 0.0582 Length: 1454  
Score: 87.00 Matches: 69  
Percent Similarity: 33.77% Conservative: 34  
Best Local Similarity: 22.62% Mismatches: 112  
Query Match: 4.32% Indels: 90  
DB: 5 Gaps: 17

US-09-763-902b-15 (1-1146) x PCT-US93-04384-2 (1-1454)

260 CTAACAGCGCTGATCTGACCAAGCCACTAATATGATGAATGAATGCTGTTTCAGG 319  
211 ValThrLeuLeuTyrSerArgSerThrAlaThrTrpGluTyrSerAlaAlaTyrAla 230  
320 CTGATGCCCTAATAATAATTTTGGACACATTCGCTTCTAGTGATTCAGGTCAGTCAATATG 379  
231 TyrGlnGlyValSerAsnPhe-----ThrTyrTyrLysLeuAsnAsn 244  
380 TTAGAT-----TGCCAACGACTTCATGATGAGCAAGCAAGCACTGATCC 418  
245 ThrAsnGlyLeuLysThrTyrGluLeuGlySgluAsp---TyrGluHisCysThrGlyTyr 263  
419 CAACACTACAGTCTTGTGTGACAGCTGGAAGTGTGAGTGAACAAACACAT 478  
264 AlaThrAsnValPheAlaProThrSerGlyTyrIleProAspGly-----Phe 280  
479 TTCTTACACAGACACTTCTGCTGACTGCTGACTCCACT----- 517  
281 SerPheAsnAsnTrpPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
518 CCCAAC-----AATACTGCTGGAACATTCGACAT----- 547  
301 ThrAsnGlnProLeuLeuIleAsnCysLeuTrpProValProSerPheGlyValAlaAla 320  
548 -----GATTGCTCCGCTTTTCAAGATTGCTAGTAAAGGGCAAAAGTCATTC 601  
321 GlnGluPheCysPheGluGluGlyAlaGlnPheSerGlnCys-----AsnGlyVal 336

602 TCATTGGTCATTAGTTCACCAATGAAATTTATGCAATTAT----- 646  
647 SerLeuAsnAsnThValAspValIleArgPheAsnLeuAsnPhetThAlaAspValGln 356  
647 SerGlyMetGlyAlaThValPheSerLeuAsnThrGlyGlyValIleLeuGluIle 376  
650 GATTGTAGACACATATATATGCTGTAACATAATTTCTTAATTTCTATCTGCG 709  
650 SerGlySerAspThValSerGluSerSer-----TyrSerGlyGly 392  
710 TCAGACCTTTTCTAGACAGCTGCCAGTTGGAGCTTGCCTCTA-----AGA 757  
710 GluIleProPheGlyIleThraspGlyProArgGlyTyrValLeuTyrAsnGlyThr 412  
758 GCTTAAACATATTTTATACATGCTTATATATACATTCACATGACATTTCTTAATA 817  
758 AlaLeuLysTyrLeuGlyThrLeuPro-----ProSerValLysGluIleAlaIle 429  
818 ATATTAAACATGATGCTGACTACATACATCTGACCTGAAACGCTATGCAAAA 877  
818 SerLysTrpGlyHisPheTyr-----IleAsnGly 439  
878 TAAATCTTTTATATATATATATATGAGATGTCAGCAATATATAC--ACTCTGGGA 934  
878 TyrAsnProPheSerThrPheProIleGlyCys--IleSerPheAsnLeuThrGlyAla 459  
935 GAGTGGAGTTTGGTTATAGTTATATTTCTAGTAAACAGCTGCTGTTTTCAG 994  
935 AlaLeuLysTyrLeuGlyThrLeuPro-----TyrThrSerTyrThrGluAlaValGlnVal 478  
995 TT--AACACT 1002  
478 AlaGluAsnThr 481

RESULT 4  
PCT-US93-04384-47  
Sequence 47, Application PC/TUS9304384  
GENERAL INFORMATION:  
APPLICANT: Miller, Timothy J.  
APPLICANT: Kieper, Sharon  
APPLICANT: Reed, Albert Paul  
APPLICANT: Jones, Elaine V.  
TITLE OF INVENTION: Compositions and Methods for Vaccination  
TITLE OF INVENTION: Against Coronaviruses  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation - Corporate  
ADDRESS: Patents  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04384  
FILING DATE: 19930507  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,171  
FILING DATE: 08-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066

FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Schreck, Patricia A.  
REGISTRATION NUMBER: 33,777  
REFERENCE/DOCKET NUMBER: SBC H85009-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US93-04384-47

Alignment Scores:  
Pred. No.: 0.0562 Length: 1454  
Score: 87.00 Matches: 69  
Percent Similarity: 33.77% Conservative: 34  
Best Local Similarity: 22.62% Mismatches: 112  
Query Match: 4.32% Indels: 90  
Gaps: 17

US-09-763-902b-15 (1-1146) x PCT-US93-04384-47 (1-1454)

260 CTAAACAGGCTGATCTGACAGACGCCACTTAATATGGAATGAAATGCTTTCAGCG 319  
211 ValThrLeuLysTyrSerArgSerSerThrAlaThrTrpGluTyrSerAlaAlaTyrAla 230  
320 CTGATGCCCTAAATATTTTGTGACACATTCCTCTAGAGAGTTCAGGATCAATG 379  
231 TyrGlnGlyValSerAsnPro-----ThrTyrTyrLysLeuAsnAsn 244  
380 TTAGAT-----TGCAACGAGTTCATGACGACCACTGATCC 418  
245 ThrAsnGlyLeuLysThrTyrGluLeuGlyCysGluAsp---TyrGluHisCysThrGlyTyr 263  
419 CAATACAGTCTCTGTTGTGACGACGACGACGACGACGACGACGACGACGACGAC 478  
264 AlaThrAsnValPheAlaProThrSerGlyTyrIleProAspGly-----Phe 280  
479 TTCTTCACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 517  
281 SerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
518 CCCAAC-----AATACGTGTGGAGATGACAGT----- 547  
301 ThrAsnGlnProLeuLeuIleAsnGlySerLeuTrpProValProSerPheGlyValAlaAla 320  
548 -----GATTGCTCCGTTTTCACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601  
321 GlnGluPheCysPheGluGlyAlaGlnPheSerGlnCys-----AsnGlyVal 336  
602 TCATTGGTCATTAGTTCACCAATGAAATTTATGCAATTAT----- 646  
647 SerLeuAsnAsnThValAspValIleArgPheAsnLeuAsnPhetThAlaAspValGln 356  
647 SerGlyMetGlyAlaThValPheSerLeuAsnThrGlyGlyValIleLeuGluIle 376  
650 GATTGTAGACACATATATATGCTGTAACATAATTTCTTAATTTCTATCTGCG 709  
650 SerGlySerAspThValSerGluSerSer-----TyrSerGlyGly 392  
710 TCAGACCTTTTCTAGACAGCTGCCAGTTGGAGCTTGCCTCTA-----AGA 757  
710 GluIleProPheGlyIleThraspGlyProArgGlyTyrValLeuTyrAsnGlyThr 412  
758 GCTTAAACATATTTTATACATGCTTATATATACATTCACATGACATTTCTTAATA 817  
758 AlaLeuLysTyrLeuGlyThrLeuPro-----ProSerValLysGluIleAlaIle 429



```

QY 818 ATATTAACACATGATCTGGTACTTAACACTACTGCAAGCCAGCCTATTGCAAAA 877
Db 430 -----SerLysTrpAlaHisPheYr-----IleasnGly 439
QY 878 TAAATCTTTTATATATATATATCTATGCGATGTCAGACAAATATAC---ACTCTGGAA 934
Db 440 TyrasnPhePheSerThrPheProIleGlyGys--IleSerPheasnLeuThrThrcGlyA 459
QY 935 GAATGAGCTTTTGTGGTTATTAGCTTAATTTCTAGTAAACACAGCTGCTGTTTCAG 994
Db 459 laserGlyAlaPheThrPheThrIleAla--TyrThrSerTyrThrGuaIalaLeuValGlnV 478
QY 995 TT---AACACT 1002
Db 478 alGluasnThr 481

RESULT 5
US-08-591-685-9
Sequence 9, Application US/08591685
Patent No. 6083733
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Thermostable xylanases
NUMBER OF SEQUENCES: 13
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/591,685
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 544 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-591-685-9

Alignment Scores:
Pred. No.: 0.0472 length: 544
Score: 86.00 Matches: 32
Percent Similarity: 38.28 Conservative: 17
Best Local Similarity: 25.00% Mismatches: 50
Query Match: 4.27% Indels: 29
DB: 3 Gaps: 4

US-09-763-902B-15 (1-1146) x US-08-591-685-9 (1-544)
QY 256 GGCACCTAACACAGGCGTATCTGGACAGGCCACCTTAATATGGAATGAAATGCTGTTTC 315
Db 281 GlyLeuIleAspGlyValcIcIyLeuGlnGlnIyHisIleasnValasp-SerProalavalIy 3000
QY 316 AGGCGTGAATGCCCTAAATATATTTTATGACATGATGCCCTCTAGAGAGTTCAGGTCACAA 3745
Db 300 scGluIleGluasPrrHlleasnLeuPheSerThrIleProGlyLeuGlnIleasn 3200
QY 376 TATGTTAGATTGGCCAAACCAAGTTCATGAGCAGCAACCACTCAGTCCCAACTACAGTTCTTGT 4355
Db 320 rGluLeuasp-----IleSerVa 326
QY 436 TGTGACCACTGGAAGTCTGCAATTTGATGGAAACAACAACATTC- 481
Db 326 lYrThrSerSerThrGlnGlnIyTyrAspThrLeuProGlnaspIleMetIleYsGlnal 346
QY 482 -----TTCAACCAAGACTTCCTGCTGACCTGCTCAGTCCACT-----CCCAACAA 5255
Db 346 aleuYsPheYsGluLeuPheGluMetLeuYsArGHisSerAspArgIleThrAsnVa 3666
QY 526 TACTGTGTGGGAAGATTGCAAGTGAATGCTTCCTCCGTTTTCAGAGATTGGTCTAGTAGTTAAAG 5855

```

```

Db      366 1ThrlleuTrpGlyLeuLysAspAsp-----TyrProTrpLeuSerLysAspAr 382
OY      586 GGGCAAAAGTCATTCATTCATT 607
      1 111 111
Db      382 gSerAsnTrpProLeuLeuPhe 389

RESULT 6
PCT-US93-04384-44
; Sequence 44, Application PC/RTUS9304384
; GENERAL INFORMATION:
; APPLICANT: Miller, Timothy J.
; APPLICANT: Kieffer, Sharon
; APPLICANT: Reed, Albert Paul
; APPLICANT: Jones, Elaine V.
; TITLE OF INVENTION: Compositions and Methods for Vaccination
; TITLE OF INVENTION: Against Coronaviruses
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SmithKline Beecham Corporation - Corporate
; ADDRESSEE: Patents
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/04384
; FILING DATE: 19930507
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/882,171
; FILING DATE: 08-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/698,927
; FILING DATE: 13-MAY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/613,066
; FILING DATE: 14-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Schreck, Patricia A.
; REGISTRATION NUMBER: 33,777
; REFERENCE/DOCKET NUMBER: SBC H85009-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 270-5015
; TELEFAX: (215) 270-5090
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US93-04384-44

Alignment Scores:
Pred. No.: 0.0784 Length: 1454
Score: 86.00 Matches: 72
Percent Similarity: 35.20% Conservative: 35
Best Local Similarity: 23.68% Mismatches: 109
Query Match: 4.27% Indels: 88
DB: 5 Gaps: 19

US-09-763-902B-15 (1-1146) x PCT-US93-04384-44 (1-1454)
OY 260 CTAACGACGGCTGTATCTGACACAGCGACCTTAATATGGAATGAATGCTGTTTCAGG 319
      ::||| ||||| ::||| ||| ::|||
Db 211 ValThrLeuLeuTyrSerArgSerSerThrAlaThrTrpCyluTyrSerAlaAlaTyrAla 230

```

```

OY 320 CTGGATGCCCTAAATAATTTTTTTTGACACTTCGCTTGTACGAGTCCAGGCTAAATG 379
Db 231 TylGInGlyValSerAsnPhe-----ThrTyrTyrLysLeuAsn 244
OY 380 TTAGAT-----TGCACACCAAGTTCATGAGCAAGCAACTCAAGTCC 418
Db 245 ThrAsnGlyLeuLysThrTyrGlnLeuGlyGlnAar---TyrGlnHisCysThrGlyTyr 263
OY 419 CAATACACAGTCTCTGTCGACCAAGTGAGCACTGGAAGTTGATGGAACCAACAACAT 478
Db 264 AlaThrAsnValAlaPheAlaIleProThrSerGlyGlyTyrIleProAspGly-----Phe 280
OY 479 TTCTTACACCAAGCACTCTCTGTCAGCTGCTCAAGTCCACT 517
Db 281 SerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSerGlyAlaGpPheVal 300
OY 518 CCCAAC-----AAATCCTGTGTGGAAGATGCAAGT 547
Db 301 ThrAsnGlnProLeuLeuLeuLeuAsnGlyLeuThrProValProSerPheGlyValAlaAla 320
OY 548 -----GATGCTCTCCGTTTTCAGATATGGCTCTACTAGTAAAGGGCAAGTCCATTC 601
Db 321 GlnGlnPheCysPheGlnGlyAlaGlnPheSerGlnCys-----SerGlyVal 336
OY 602 TCATTTGGTGCATTAAGTTCACACCAATGGAATTTATGCAATTAT 646
Db 337 SerLeuAsnAsnThrValAlaIleAlaGpPheAsnLeuAsnPheThrAlaAspValGln 356
OY 647 -----TTGATGTTAGACAGCT 667
Db 357 SerGlyMetCysAlaThrValPheSerLeuAsnThrThrGlnGlyValIleLeuGlnVal 376
OY 668 ATATGT-----GCTGAACATAATTTCTTAATATTTTCTATTCCTGTCA 712
Db 377 SerCysTyrAsnAspThrValSerGlnSerSerPhe-----TyrSerTyrGlyGln 393
OY 713 GCACCTTTCTAGCAGCTGCCAGTTTGAGAGCATTCCTCTTA-----AGAGCT 766
Db 394 IleProPheGlyIleThrAspGlyProAlaGlyTyrCysTyrValLeuTyrAsnGlyThrAla 413
OY 761 TTAACACTATTTTATACATGCCCTATATATACATTCACATAAGACATCTTATATATA 820
Db 414 LeuLysTyrLeuGlnIleThrLeuPro-----ProSerValLysGlnIleAlaIle----- 428
OY 821 TTAACACATGATCTTGCTACTACATACTCACTGACACCAAGCCAGCTATTGCAAAATAA 880
Db 430 -----SerLysThrGlyHisPheTyr-----IleAsnGlyTyr 440
OY 881 AATCTTTTATATATATATATATATATATATATATATATATATATATATATATATATAT 937
Db 441 AsnPhePheSerThrPheProIleAspCys---IleSerPheAsnLeuThrThrLysAsp 466
OY 938 GTGAGATTTTGTGTTATAGTGAATATTTTCTAGTAAACACAGCTGCCTGTTTCAGT- 996
Db 460 ergLysAlaMetThrPheIleAla---TyrThrSerTyrThrGlnAlaLeuValGlnValG 479
OY 997 --AAACT 1002
Db 479 LysAsnThr 481

RESULT 7
PCT-US93-04384-45
; Sequence 45, Application: PC/TUS9304384
; GENERAL INFORMATION:
; APPLICANT: Miller, Timothy J.
; APPLICANT: Klepfer, Sharon
; APPLICANT: Reed, Albert P.
; APPLICANT: Jones, Elaine V.
; TITLE OF INVENTION: Compositions and Methods for Vaccination
; TITLE OF INVENTION: Against CoronaviLuses
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation - Corporate

```

ADDRESS: Patents  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04384  
FILING DATE: 19930507  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,171  
FILING DATE: 08-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Schreck, Patricia A.  
REGISTRATION NUMBER: 33,777  
REFERENCE/DOCKET NUMBER: SBC HB5009-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US93-04384-45

	Alignment Scores:			
Pred. No.:	0.0784	Length:	1454	
Score:	86.00	Matches:	72	
Percent Similarity:	35.20%	Conservative:	35	
Best Local Similarity:	23.68%	Mismatches:	109	
Query Match:	4.27%	Indels:	88	
DB:	5	Gaps:	19	

US-09-763-902B-15 (1-1146) x PCT-US93-04384-45 (1-1454)

QY	260	CTAACGAGCGTGTATCTGTGACACAGGCCACTTAATAATGGAAATGCAATGCTGTTTCAGGG	319
		:::::	
Db	211	VAlTrLeuLeuLysSerArgSerSerThrAlaThrThrGluLutrySerAlaAlaLaTyraLa	230
		:::::	
QY	320	CTGATGCGCCCTMAAATAATTTTTGTGCACACATTGCCCTTAGTGAGTTCACAGCTCAATATG	379
		:::::	
Db	231	TyrgInglYvalSerAsnPh-----ThrrTyrrLySleuAsnAsn	244
		:::::	
QY	380	TTAGAT-----TCCACACCGTTCAATGACGACGAACAATCAGTCC	418
		:::::	
Db	245	ThraSnglYleuLYstHrrTYrgIuLeucYsgluAsP--TyrdIuhIsCYstrHrgLYtyr	263
		:::::	
QY	419	CAAACTGACGTCTTGTTGTGACACAGTGAACACTGTGAATTTGATGGAAACAACAACAT	478
		:::::	
Db	264	AlathrasnaValphalaProthrsergLYtyrIIleProaspoly-----Phe	280
		:::::	
QY	479	TTCTTCAACGACAACTCTCGCTGACTGCTCAGTCACCT-----	517
		:::::	
Db	281	SerPheaSnAsnUtrPheLeuLeuThraSnsErSerThrrPheValsercylarPheval	300
		:::::	
QY	518	CCCAC-----AATACTGNGTGAAGATTGCAAT-----	547
		:::::	
Db	301	ThraSnglPrOleuLeuIleasnCySleutrProvalProSerPheclYvalAlaAla	320



Db 430 -----SerLysTrpGlyHisPheTyr-----IleAsnGly 439  
QY 878 TAAATCTTTTAAATATATATCTATGGATGACACAAATATAAC---ACTCTGGGAA 934  
Db 440 TyrAsnPhenPheSerThrPheProIleGlyCys--IleSerPheAsnLeuThrGly 459  
QY 935 GAATGAGGTTTGTGTTATAGTTATTTCTAGTAAACAGCTGCTGTTTCAG 994  
Db 459 alSerGlyAlaPheTrpThrIleAla--TyrThrSerTyrThrGluAlaLeuValGln 478  
QY 995 TT---AACACT 1002  
Db 478 alGluAsnThr 481  
RESULT 9  
PCT-US91-08525-24  
Sequence 24, Application PC/TUS9108525  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/08525  
FILING DATE: 19911114  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-08525-24  
Alignment Scores:  
Pred. No.: 0.0749 Length: 748  
Score: 85.00 Matches: 69  
Percent Similarity: 33.77% Conservative: 34  
Best Local Similarity: 22.62% Mismatches: 112  
Query Match: 4.22% Indels: 90  
Gaps: 17  
US-09-763-902b-15 (1-1146) x PCT-US91-08525-24 (1-748)  
QY 260 CTACACGAGCTGATCGACACAGCCACTATATATGAAATGAAATGCTGTTTCAGG 319  
Db 211 ValThrLeuLeuTyrSerArgSerSerThrAlaThrTrpGluTyrSerAlaAlaTyrAla 230

QY 320 CTGATGCCCTAAATATATTTTGGACATTCCTCTACTGAGTCCAGCTCAATATG 379  
Db 231 TyrGlnGlyValSerAsnPhe-----ThrTyrTyrLeuAsnAsn 244  
QY 380 TTACAT-----TGCCACACAGTTCATGTAGCAACACACAGTC 418  
Db 245 ThrAsnGlyLeuLysThrTyrGluLeuGlyGlnLysPhe---TyrGluHisCysTyrGlyTyr 263  
QY 419 CAACACAGCTGCTGTTGTTGATGACAGCTGAGAGCTTGATGGAACAAACACAT 478  
Db 264 AlaThrAsnValPheAlaProThrSerGlyLysTyrIleProAspGly-----Phe 280  
QY 479 TTCTTCACACAGAACTTCCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 517  
Db 281 SerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
QY 518 CCCAAC-----AATACGTGTGGAAGATTCGACGT----- 547  
Db 301 ThrAsnGlnProLeuLeuIleAsnCysLeuTrpProValProSerPheGlyValAlaAla 320  
QY 548 -----GATTCCTTCCTGTTTCAGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 601  
Db 321 GlnGluPheCysPheGlnGlyAlaGlnPheSerGlnCys-----AsnGlyVal 336  
QY 602 TCATTTGGTCCATTAGTCCAGCAATTCGAATTTATGGAATATAT----- 646  
Db 337 SerLeuAsnAsnThrValAlaPheValIleArgPheAsnLeuAsnPheThrAlaAspValGln 356  
QY 647 -----TTT 649  
Db 357 SerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyGlyValIleLeuGluIle 376  
QY 650 GATGTGAGACCATATATATGTCGCAACCTAAATTTCTTATATTTCTATCTGCTG 709  
Db 377 SerCysTyrSerArgPheValSerGlnSerSerSer-----TyrSerTyrGly 392  
QY 710 TCACACACCTTTTCTAGACAGTCCAGCTTGGAGCATTCGCCCTCA-----AGA 757  
Db 393 GluIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeuTyrAsnGlyThr 412  
QY 758 GCTTAAACATATTTTTCATACCTGCTTATATACATTCCTCAATGACATTCCTTATATA 817  
Db 413 AlaLeuLysTyrLeuGlyThrLeuPro-----ProSerValLysGluIleAlaIle 429  
QY 818 ATATTAAACATATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 877  
Db 430 -----SerLysTrpGlyHisPheTyr-----IleAsnGly 439  
QY 878 TAAATCTTTTAAATATATATCTATGGATGACACAAATATAAC---ACTCTGGGAA 934  
Db 440 TyrAsnPhenPheSerThrPheProIleGlyCys--IleSerPheAsnLeuThrGly 459  
QY 935 GAATGAGGTTTGTGTTATAGTTATTTCTAGTAAACAGCTGCTGTTTCAG 994  
Db 459 alSerGlyAlaPheTrpThrIleAla--TyrThrSerTyrThrGluAlaLeuValGln 478  
QY 995 TT---AACACT 1002  
Db 478 alGluAsnThr 481  
RESULT 10  
PCT-US93-04384-4  
Sequence 4, Application PC/TUS9304384  
GENERAL INFORMATION:  
APPLICANT: Miller, Timothy J.  
APPLICANT: Klepfer, Sharon  
APPLICANT: Reed, Elaine V.  
APPLICANT: Jones, Albert P.  
TITLE OF INVENTION: Compositions and Methods for Vaccination  
TITLE OF INVENTION: Against Coronaviruses  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation - Corporate





Db 297 GlyArgPheValThrAsnGlnProLeuValAlaAsnCysLeuThrProValProSerPhe 316  
QY 548 -----GATGCTTCCCTTTTCAAGATTGGCTCTGACTTAAGGGC 589  
Db 317 GlyValAlaIaGlnGluPheCysPheGluGlyAlaGlnPheSerGlnCys 333  
QY 590 AAAAGTCCTTCTCATTTGGTCCATTAGTCCAGCAATGAATTATGCAATTAT 646  
Db 334 ---AsnGlyValSerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnPheThr 352  
QY 646 ----- 646  
Db 353 AlaAspValGlnSerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyVal 372  
QY 647 -----TTGATGTAGAGAGCACTAATATATGCTGAAACTAATTTCTTAATAT 697  
Db 373 IleLeuGluIleSerCysTyrSerAspThrValSerGlnUserSerSer 388  
QY 698 TTTCTATTCTGTCAGCACCCTTTCTAGACAGCTCCAGTTGAGCATTCCTCTCA 754  
Db 389 TyrSerTyrGlyGluIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeu 408  
QY 755 -----AGACTTAAACATAATTTTATACATGCTATATACATTCACATAATGAC 805  
Db 409 TyrAsnGlyThrAlaLeuIleGlyThrLeuPro-----ProSerValGlyGlu 426  
QY 806 ATCTCTATATATATATATTAACACATGATCTGTGTAACATCACTCACTGAAACCCAGC 865  
Db 427 IleAlaIle-----SerTyrIlePheIlePheIle 436  
QY 866 CTATTGCAAAAAATCTTTTATATATATCTATCTGAGTCTGACCAATATATAC 924  
Db 437 ---IleAsnGlyTyrAsnPhePheSerThrPheProIleGlyCys--IleSerPheAsn 455  
QY 925 ---ACTGGGAGAGAGAGAGTGTGTGTTAGTAAATTTCTAGTAAACACAGT 982  
Db 455 eutThrGlyAlaSerGlyAlaPheThrPheIleAla---TyrThrSerTyrThrGlu 474  
QY 983 GCCTGTTTCAGTT---ACACT 1002  
Db 474 IalLeuValGlnValGluAsnThr 481

RESULT 13  
PCT-US93-04384-46  
Sequence 46, Application PC/TUS9304384  
GENERAL INFORMATION:  
APPLICANT: Miller, Timothy J.  
APPLICANT: Klepfer, Sharon  
APPLICANT: Reed, Albert Paul  
APPLICANT: Jones, Elaine V.  
TITLE OF INVENTION: Compositions and Methods for Vaccination  
TITLE OF INVENTION: Against Coronaviruses  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation - Corporate  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04384  
FILING DATE: 19930507  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,171

FILING DATE: 08-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Schreck, Patricia A.  
REGISTRATION NUMBER: 33,777  
REFERENCE/DOCKET NUMBER: SBC H85009-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US93-04384-46

Alignment Scores:  
Pred. No.: 0.403 Length: 1454  
Score: 80.50 Matches: 74  
Percent Similarity: 32.83% Conservative: 34  
Best Local Similarity: 22.49% Mismatches: 124  
Query Match: 4.00% Indels: 97  
Gaps: 18

US-09-763-902b-15 (1-1146) x PCT-US93-04384-46 (1-1454)

QY 209 GAGAGGTTGTCATATTTAC-----TATGAGACAAATGATATA 247  
Db 187 AspIlePheValThrAlaIleThrIleSerGlyValArgSerTyrAsnThrPheAsnAsn 206  
QY 248 AGAAGAGGGGAGGAGTACAGGCTGTATCTGACAGGACCACTTAATATGGAATGAAAT 307  
Db 207 TrpPheAsnAsnValThrLeuLeuTyrSerArgSerThrAlaThrIleGluTyrSer 226  
QY 308 GCTGTTTCAGGGCTGATGCTCCCTAATATATTT-----TTGAC 346  
Db 227 AlaAlaIleValTyrGlnGlyValSerAsnPheThrTyrTyrTyrLeuAsnAsnThrAsn 246  
QY 347 ACATTGCTTCTAGTGAATTCAGTCAATATGTTGACCAACAGTTCATGACCA 406  
Db 247 GlyLeuTyrThrTyrGluPhe-----CysGluAsp---TyrGluTyr 259  
QY 407 GCAACTCAGTCCCAACACTAGTCTGTGTGACACAGTGAAGTGAAGTTGATGCA 466  
Db 260 CysThrGlyTyrAlaThrAsnValPheAlaProThrValGlyTyrIleProAspGly 279  
QY 467 AACAAACAACATTTCTTACACAGACTCTGCTGACTGCTCACTGCTCACT 517  
Db 280 -----PheSerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSer 296  
QY 518 -----CCCAAC-----AATATGCTGGAAGATTCACAGT--- 547  
Db 297 GlyArgPheValThrAsnGlnProLeuValAlaCysLeuThrProValProSerPhe 316  
QY 548 -----GATGCTTCCGTTTTCAGAGTGTGCTAGTATTAAGGGGC 589  
Db 317 GlyValAlaIaGlnGluPheCysPheGluGlyAlaGlnPheSerGlnCys 333  
QY 590 AAAAGTCCTTCTCATTTGGTCCATTAGTCCAGCAATGAATTATGCAATTAT 646  
Db 334 ---AsnGlyValSerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnPheThr 352  
QY 646 ----- 646  
Db 353 AlaAspValGlnSerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyVal 372  
QY 647 -----TTGATGTAGAGAGCACTAATATATGCTGAAACTAATTTCTTAATAT 697

PRIOR APPLICATION DATA.

RESULT 15  
US-08-392-459-32  
Sequence 32, Application US/08392459  
Patent No. 6280974  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S  
PROTEINS  
TITLE OF INVENTION: Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/392,459  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927



FILING DATE: 13-MAY-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/613,066  
 FILING DATE: 14-NOV-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: King, William T.  
 REGISTRATION NUMBER: 30,954  
 REFERENCE/DOCKET NUMBER: SBC 14532B  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 270-5015  
 TELEFAX: (215) 270-5090  
 INFORMATION FOR SEQ ID NO: 32:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1454 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-392-459-32

Alignment Scores:  
 Pred. No.: 0.543 Length: 1454  
 Score: 79.50 Matches: 77  
 Percent Similarity: 34.15% Conservative: 35  
 Best Local Similarity: 23.48% Mismatches: 121  
 Query Match: 3.95% Indels: 95  
 DB: 4 Gaps: 20

US-09-763-902b-15 (1-1146) x US-08-392-459-32 (1-1454)

```

OY 209 GAGGAGTTTGTCATATTTC-----TATGAGACAATGATATA 247
Db 187 AspGluPheValThrAlaIleSerGlyArgSerTyrAsnThrPasnIleAsnAsn 206
OY 248 AGAAGAGGCGCCTACACAGGCTGATCTGACAGGACCTTAATGGAATGAAAT 307
Db 207 TrpPheAsnAsnValThrLeuLeuTyrSerArgSerThrAlaThrIlePglutyrSer 226
OY 308 GCTGCTTCAGGCGCTGATGCCCTAATATTT-----TTTGAC 346
Db 227 AlaAlaIleTyrValTyrGlnGlyValSerAsnIleThrTyrLysLeuAsnAsnThrAsn 246
OY 347 ACATTCGCTTCTAGTACGATTCACAGTCAATATGTTAGATTCACACAGGCTGATGACCA 406
Db 247 GlyLeuLeuThrTyrGluPhe-----CysGluAsp---TyrGluTyr 259
OY 407 GCAACTCAGTCCCAACAGTCTGTTGTGACACAGTGAAGTGAAGTTGATGA 466
Db 260 CysThrGlyTyrAlaThrAsnValPheAlaProThrValGlyGlyTyrIleProAspGly 279
OY 467 AACAAACAATTTCTTCACACAGACTCTGCTGACTGCTGACTGCTGACTGCTGACTGCTGACT 517
Db 280 -----PheSerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSer 296
OY 518 -----CCCAC-----AATACGTGTGTGAGATTCGCACT--- 547
Db 297 GlyArgPheValThrAsnGlnProLeuLeuValAsnGlyLeuTyrProValProSerPhe 316
OY 548 -----GATTGCTTCGTTTCAAGATTCGCTAGTAGTAGTAAAGGGCG 589
Db 317 GlyValAlaAlaGlnGluPheCysPheGlnGlyAlaGlnPheSerGlnCys----- 333
OY 590 AAAAGTCCATTTCATTGTCATTTAGTCCAGCAATGGAATTTATGTAATAT 646
Db 334 ---SerGlyValSerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnIlePheThr 352
OY 647 -----TTGTGTGTAGAGCACT----- 664
Db 353 AlaAspValGlnSerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyGlyVal 372
OY 665 -----ATAATATGT-----GCTGAACAATAATTTCTTAATATATTT 700
Db 373 IleLeuGluValSerCysTyrAsnAspThrValSerGluSerSerPhe-----Tyr 389
  
```

```

OY 701 CTATTCCTGTCAGCACCTTTCTGACAGCTGCCAGTTTGAGACATGCGCTCTA----- 754
Db 390 SerTyrGlyGluIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeuTyr 409
OY 755 -----AGAGCTTAAACATATTTTTCATGCTTATATATATATATATATATATATATATAT 808
Db 410 AsnGlyThrAlaLeuTyrLeuGlyThrLeuPro-----ProSerValIleGluIle 427
OY 809 CTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 868
Db 428 AlaIle-----SerLysTrpGlyHisPheTyr----- 436
OY 869 TTGCAAAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 925
Db 437 IleAsnGlyTyrAsnIlePhePheSerThrPheProIleAspCys---IleSerPheAsnLeuT 456
OY 926 CTTGCGAAGAGTGAAGTATTTTGTATATATATATATATATATATATATATATATATATATATAT 985
Db 456 hrThrGlyAspSerGlyAlaPheThrIleAla---TyrThrSerTyrThrGluAla 475
OY 986 TGTTCAGTT---AACACT 1002
Db 475 euValGlnValGluAsnThr 481
  
```

Search completed: June 24, 2003, 19:40:32  
 Job time : 32.789 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:20:48 ; Search time 54,8062 Seconds  
(without alignments)  
5572.555 Million cell updates/sec

Title: US-09-763-902b-15  
Perfect score: 2013  
Sequence: 1 taccctcgsgagagaatggyg.....tttaagratattttgtttt 1146

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL-frame-n2p.model -DEV-nlp  
-Q-cg2.1/USPTO-SPOOL/US09763902/runat\_24062003.130340.12405/app\_query.fasta\_1.2254  
-DB-A.Geneseq\_101002 -OPMT-fastaan -SUFFIX-n2p.rag -MINMATCH-0.1 -LOOPTOL-0  
-LOOPTXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-biosum62 -TRANS-human40.cdl  
-LIST-45 -DOCALLIGN-200 -THR SCORE-pct -THR MAX-100 -THR MIN-0 -ALIGN-15  
-MODE-LOCAL -OUTFMT-pct -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000  
-USER-US09763902-EGCN\_L1-1.162-@runat\_24062003.130340.12405 -NCPN-6 -ICPU-3  
-NO\_MAP -LANGQUERY-NEG-SCORES-30 -WAIT -DSPBLOCK-100 -IONOLOG  
-DEV-TIMEOUT-120 -WARN-TIMEOUT-30 -THREDS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6  
-FGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database :

A.Geneseq\_101002.\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	874	43.4	197	22	AB95379
2	749	37.2	142	21	AA982323
3	749	37.2	142	23	AB804459
4	727	36.1	162	22	AA016227
5	584.5	29.0	164	22	AA016228
6	584	29.0	140	21	AA082318
7	310	15.4	133	22	AB865017
8	114	5.7	688	22	AB867404
9	114	5.7	680	22	AB866052
10	113.5	5.6	125	21	AA012073
11	113.5	5.6	153	21	AA005145
12	112.5	5.6	123	21	AA008682
13	112.5	5.6	126	21	AA008681
14	112.5	5.6	131	21	AA005147
15	112.5	5.6	134	21	AA005146
16	105	5.2	122	21	AA005146
17	105	5.2	157	21	AA029867
18	102.5	5.1	130	22	AB871855
19	91.5	4.5	170	22	AA025600
20	90.5	4.5	428	21	AA021068
21	90.5	4.5	436	21	AA021067
22	89	4.4	472	21	AA046342
23	88.5	4.4	143	21	AA054152
24	87.5	4.3	460	21	AA046343
25	87.5	4.3	699	14	AA030512
26	87	4.3	1454	13	AA024395
27	87	4.3	1454	14	AA042464
28	87	4.3	1454	14	AA042477
29	86	4.3	471	21	AA020809
30	86	4.3	544	17	AA087014
31	86	4.3	703	22	AB0113036
32	86	4.3	1017	22	AB008678
33	86	4.3	1454	14	AA042473
34	86	4.3	1454	14	AA042474
35	85	4.2	1594	9	AA081183
36	84.5	4.2	459	21	AB020810
37	84	4.2	938	22	AB010442
38	83.5	4.1	1127	22	AA033877
39	83.5	4.1	1158	22	AA036929
40	83	4.1	284	20	AA035356
41	82	4.1	746	23	AA074631
42	82	4.1	752	22	AA07205
43	82	4.1	755	22	AA07206
44	82	4.1	755	22	AB061889
45	82	4.1	755	23	AA051198

## ALIGNMENTS

RESULT 1  
ID AAB95379 standard; Protein; 197 AA.  
AC AAB95379;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:17707.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI: 2001-318749/34..  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8: SEQ ID 17707; 2537bp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification, where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 CC  
 SQ Sequence 197 AA:

Alignment Scores:  
 Pred. No.: 4,19e-98 Length: 197  
 Score: 874.00 Matches: 167  
 Percent Similarity: 98.82% Conservative: 0  
 Best Local Similarity: 98.82% Mismatches: 2  
 Query Match: 43.42% Indels: 0  
 DB: 22 Gaps: 0

US-09-763-902B-15 (1-1146) x AAB95379 (1-197)

OY 74 TCATCCGAGCCCGGACCGGACGACGTAGAGAGTAGTACGCCGACATGCGAGAA 133  
 DB |||||  
 DB 29 SerHisSerSerProAlaAspArgThrArgGluGluValThrProThrLeuProGlu 48  
 OY 134 CACACTGCTACAGGTCCAGATGCCAGCTGTGATTTAAACTATATGATGATGAG 193  
 DB |||||  
 DB 49 HisThrAlaThrArgSerGlnMetAlaThrSerLeuAspPheLysThrTyrAlaAspGln 68  
 OY 194 GCATGTAGAGCTGTGAGAGAGTTTGTCAATATATGATGACACATGATTAAGAAGA 253  
 DB |||||  
 DB 69 AlaCysArgAlaAlaGluIlePheValAsnIleTyrTyrGluThrMetAspLysArgArg 88  
 OY 254 CGGGACACTACCAAGCTGTATCTGACAGCAAGCCACCTTAATATGAAAGGAATCTGTT 313  
 DB |||||  
 DB 89 ArgAlaLeuThrThrArgLeuTyrLeuAspLysAlaThrIleLeuTyrAsnGlnLysAlaVal 108  
 OY 314 TCAGGCTGATGCCCTAAATATTTTTCACACATTCCTTCTAGTGAAGTTCAGATC 373

DB |||||  
 DB 109 SerGlyLeuAspAlaLeuAsnAsnPhePheAspThrLeuProSerSerGluPheGlnVal 128  
 OY 374 AATATGTAGATTGGCAACCAAGTTCATAGACAGCAATCTGATCCCAACATCACTTTT 433  
 DB |||||  
 DB 129 AsnMetLeuAspCysGlnProValHisGluIleAlaThrGlnSerGlnThrThrValLeu 148  
 OY 434 GTTGACCACTGAGCACTGTGATGATGGAACAACAACAATTTCTTCAATCAGAAC 493  
 DB |||||  
 DB 149 ValValThrSerGlyThrValLysPheAspLysGlnHisPheAsnGlnAsn 168  
 OY 494 TTCTGTGACTGCTCACTCCACATCCACAACTACTGTGTGAGAGATGCAAGTATTC 553  
 DB |||||  
 DB 169 PheLeuLeuThrAlaGlnSerThrProAsnAsnThrValTyrPheAlaSerAspCys 188  
 OY 554 TTCGTTTCAAGATGTGCTAGTAGT 580  
 DB |||||  
 DB 189 PheArgPheGlnAspTyrPheSerSer 197

# RESULT 2

AAH82323

ID AAH82323 standard; Protein: 142 AA.

XX AAH82323;

AC 19-JUN-2000 (first entry)

DE Human protein transport molecule (PTM) SEQ ID NO:7:

XX Human: protein transport molecule; PTM; diagnosis; cytostatic;  
 KW antihypertic; antidiabetic; immunosuppressant; antileukocytic;  
 KW antiallergic; antidiabetic; antileukemic; antileukemic; osteopathic;  
 KW dermatological; antianemic; antipruritic; hepatotropic; antipruritic;  
 KW antileukemic; antileukemic; antileukemic; antileukemic; antileukemic;  
 KW immune disorder; cell proliferative disorder; secretory disorder;  
 KW urticaria; allergy; abnormal vesicle trafficking; asthma;  
 KW autoimmune haemolytic anaemia.

OS Homo sapiens.

PN WO200012703-A2.

PD 09-MAR-2000.

XX 26-AUG-1999; 99WO-US19616.

PR 27-AUG-1998; 98US-0098206.

PA (INCY-) INCYTE PHARM INC.

XX Tang YF, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;

PI Gorgone GA, Baughn MR, Patterson C;

DR WPI: 2000-256642/22.

DR N-PSDB: AAA08041.

PT New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis

PS Claim 1; Page 68; 75pp; English.

CC AAA08035 to AAA08042 encode the human protein transport-associated  
 CC molecules (PTMs) given in AAH82317 to AAH82324. The PTMs have  
 CC cytosolic, antileukemic, antileukemic, immunosuppressant,  
 CC antileukemic, antileukemic, antileukemic, antileukemic, antileukemic,  
 CC antileukemic, osteopathic, dermatological, antileukemic, antileukemic,  
 CC hepatotropic, antileukemic, antileukemic, antileukemic, antileukemic,  
 CC regulate protein transport. PTM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PTM. PTM polynucleotides are useful for  
 CC diagnosing conditions associated with PTM, comprising detecting PTM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the

PD		07-NOV-2001.	(first entry)
19-SEP-2001.			

AA	07-NOV-2001	(first entry)
DT		

XX Human novel secreted protein, Seg ID 1180.  
DE  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic;  
KW cyostatic; cardiant; vasotropic; cerebroprotective; noctopic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnery; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
XX Homo sapiens.  
XX  
XX W0200155322-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01341.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184654.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225265.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226868.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 05-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232060.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.



PR	18- AUG -2000	2000US -02266781
PR	22- AUG -2000	2000US -02266681
PR	22- AUG -2000	2000US -02266688
PR	22- AUG -2000	2000US -02271882
PR	23- AUG -2000	2000US -02270709
PR	30- AUG -2000	2000US -02298924
PR	01- SEP -2000	2000US -02292887
PR	01- SEP -2000	2000US -02292943
PR	01- SEP -2000	2000US -02293445
PR	01- SEP -2000	2000US -02293444
PR	05- SEP -2000	2000US -02295109
PR	05- SEP -2000	2000US -02295133
PR	06- SEP -2000	2000US -0230437
PR	06- SEP -2000	2000US -0230438
PR	08- SEP -2000	2000US -02311422
PR	08- SEP -2000	2000US -02311443
PR	08- SEP -2000	2000US -02311444
PR	08- SEP -2000	2000US -02311413
PR	08- SEP -2000	2000US -02311410
PR	14- SEP -2000	2000US -02324010
PR	14- SEP -2000	2000US -02330663
PR	14- SEP -2000	2000US -02330664
PR	14- SEP -2000	2000US -02330655
PR	14- SEP -2000	2000US -02330656
PR	21- SEP -2000	2000US -02344273
PR	21- SEP -2000	2000US -02344274
PR	25- SEP -2000	2000US -0234997
PR	25- SEP -2000	2000US -0234998
PR	25- SEP -2000	2000US -02358584
PR	27- SEP -2000	2000US -02358583
PR	27- SEP -2000	2000US -02358536
PR	29- SEP -2000	2000US -02363527
PR	29- SEP -2000	2000US -02363567
PR	29- SEP -2000	2000US -02363568
PR	29- SEP -2000	2000US -02363570
PR	02- OCT -2000	2000US -02368602
PR	02- OCT -2000	2000US -02370337
PR	02- OCT -2000	2000US -02370338
PR	02- OCT -2000	2000US -02370339
PR	02- OCT -2000	2000US -02370400
PR	13- OCT -2000	2000US -02393935
PR	13- OCT -2000	2000US -02393937
PR	20- OCT -2000	2000US -02409620
PR	20- OCT -2000	2000US -02411261
PR	20- OCT -2000	2000US -02411785
PR	20- OCT -2000	2000US -0241186
PR	20- OCT -2000	2000US -0241187
PR	20- OCT -2000	2000US -0241808
PR	20- OCT -2000	2000US -0241809
PR	08- NOV -2000	2000US -02464777
PR	08- NOV -2000	2000US -02464788
PR	08- NOV -2000	2000US -02465523
PR	08- NOV -2000	2000US -02465524
PR	08- NOV -2000	2000US -02465525
PR	08- NOV -2000	2000US -02465526
PR	08- NOV -2000	2000US -02465527
PR	08- NOV -2000	2000US -02465528
PR	08- NOV -2000	2000US -02465529
PR	08- NOV -2000	2000US -02466609
PR	08- NOV -2000	2000US -02466610
PR	08- NOV -2000	2000US -02466611

PR 08-NOV-2000; 2000US-02466513.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249255.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249309.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251980.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251866.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251988.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
DR WPI: 2001-488783/53.  
DR N-PSDB; AAS26215.  
XX  
XX  
PI New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating, or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX  
XX  
PS Claim 11; SEQ ID NO 1181; 980pp; English.  
XX  
XX  
CC The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunoabsorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention.

Pred. No.: 1,47e-62 Length: 164  
Score: 584.50 Matches: 106  
Percent Similarity: 90.85% Conservative: 23  
Best Local Similarity: 74.65% Mismatches: 12  
Query Match: 29.04% Indels: 1  
DB: 22 Gaps: 1

US-09-763-902b-15 (1-1146) x AAU16228 (1-164)

QY 152 CAGATGGCCAGCTCTGATTTAAACTATATAGATCGCATGTAGAGCTGCTGAG 211  
D 24 Glnuetaia---ServalaspheleysthryValaspGlnalacysargalalaia 42  
QY 212 GAGTTTGCATATTTACTATGACATGATGATAAGAGAGCGGACCTACACAGGCTG 271  
D 43 GlupheValasnvalITTYTThThmetaspysargatgaleuSeatrgleu 62  
QY 272 TATCTGACAGAGCCACCTAATATGAAATGAGTGTTCAGGGCTGATGCCCTA 331  
D 63 TyrmctgIYThralatThrleuValtrpansnGlnalalaserGlyGlnGlnSerleu 82  
QY 332 AATAATTTTGTGACATATGCTTCTAGTATGCCAGTCATATTTGATTTGCCAA 391  
D 83 SerGlnpHepeglnuMetleuproSerSerGlnpHeGlnIleSerValalaspGln 102  
QY 392 CCACTTCATGAGCAAGACATGATCCCAACTACATGCTTGTTCAGACAGGAGACT 451  
D 103 ProValHisaspGlnalalathrProserGlnThThrValleuValalalacysGln 122  
QY 452 GTGAGTTTGTATGAAACAACAACATTTCTTCAACGAGACTTCTGCTGCTGAG 511  
D 123 ValyspHeGlnuGlnalasnGlnalargaspheasnGlnanpHeleuThralacIn 142  
QY 512 TCCACTCCCAACAATGCTGTGTGAGATTCGAAAGTATGCTTCCGTTTCAGATTGG 571  
D 143 AlaserProserAanThValtrpLysIlealaserApsCyspHeatrgpHeGlnaspTrp 162  
QY 572 TCTAGT 577  
D 163 Alaser 164

RESULT 6  
ID AAY82318 standard; Protein: 140 AA.

XX AAY82318;  
AC 19-JUN-2000 (first entry)  
XX 19-JUN-2000 (first entry)  
DE Human protein transport molecule (PRAM). SEQ ID NO:2.  
XX Human protein transport molecule (PRAM). SEQ ID NO:2.  
XX Human, protein transport molecule; PRAM; diagnosis; cytosolic;  
XX antihistaminic; antidiabetic; immunosuppressant; antihistaminic;  
XX antiallergic; antidiabetic; antihistaminic; antihistaminic;  
XX dermatological; antineoplastic; antiproliferative; hepatotropic; antitumor;  
XX antineoplastic; antihistaminic; antiproliferative; hepatotropic; antitumor;  
XX immune disorder; cell proliferative disorder; secretory disorder;  
XX urticaria; allergy; abnormal vesicle trafficking; asthma;  
XX autoimmun haemolytic anaemia.  
XX Homo sapiens.  
XX OS  
XX WO200012703-A2.  
XX 09-MAR-2000.  
XX 26-AUG-1999; 99WO-US19616.  
XX 27-AUG-1998; 98US-0098206.  
XX (INCYTE PHARM INC.  
XX Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ.

PI Gorgone GA, Baughn MR, Patterson C;  
XX WPI: 2000-256642/22.  
DR N-PSDB: AAA08036.  
XX New human protein transport-associated polypeptide and polynucleotide  
XX useful for diagnosis, prevention and treatment of cell proliferative  
XX and secretory disorders such as leukemia, cystic fibrosis  
XX Claim 1; Page 61-62; 75BP; English.  
XX AAA08035 to AAA08042 encode the human protein transport-associated  
XX molecules (PRAMs) given in AAY82318 to AAY82324. The PRAMs have  
XX cytosolic, antihistaminic, antidiabetic, immunosuppressant,  
XX antihistaminic, antidiabetic, antidiabetic, antidiabetic,  
XX antihistaminic, antidiabetic, antidiabetic, antidiabetic,  
XX hepatotropic, antitumor, antineoplastic, antineoplastic, and  
XX regulating protein transport. PRAM proteins and antagonists are useful for  
XX preventing or treating a disorder associated with decreased or increased  
XX expression or activity of PRAM. PRAM polynucleotides are useful for  
XX diagnosing conditions associated with PRAM, comprising detecting PRAM by  
XX forming a hybridisation complex, preferably after PCR amplifying the  
XX biological sample. Diseases prevented, treated or diagnosed include cell  
XX proliferative disorders such as cancers, immune disorders, secretory  
XX disorders and other conditions associated with abnormal vesicle  
XX trafficking, such as allergies, asthma, urticaria and autoimmune  
XX haemolytic anaemia. Anti-PRAM antibodies may be used as antagonists, as  
XX a targeting or delivery mechanism for bringing pharmaceutical agents  
XX into cells or tissues expressing PRAM and for diagnosis of PRAM-related  
XX disorders. PRAM, its catalytic or immunogenic fragments are useful for  
XX drug screening using libraries of compounds. PRAM polynucleotides are  
XX useful for generating hybridisation probes useful in mapping the  
XX naturally occurring genomic sequences.  
XX Sequence 140 AA:  
SQ

Alignment Scores:  
Pred. No.: 1,58e-62 Length: 140  
Score: 584.00 Matches: 104  
Percent Similarity: 91.30% Conservative: 22  
Best Local Similarity: 75.36% Mismatches: 12  
Query Match: 29.01% Indels: 0  
DB: 21 Gaps: 0

US-09-763-902b-15 (1-1146) x AAY82318 (1-140)

QY 164 TCTGTGATTTAAACTATATGATGATGATGATGATGATGATGATGATGATGAT 223  
D 3 ServalaspheleysthryValaspGlnalacysargalalaiaGlnuGlnpHealasn 22  
QY 224 ATTTACTGTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 283  
D 23 ValTYTThThThmetaspysargatgaleuSeatrgleuTmetolYThr 42  
QY 284 GCCACTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 343  
D 43 AlathrleuValtrpansnGlnalalaserGlyGlnGlnSerleuSeatrgleu 62  
QY 344 GACACATGCTTCTGAGTGTTCAGGCTCAATATGATGATGATGATGATGATGATGAT 403  
D 63 GlnuetaiaProserSerGlnpHeGlnIleSerValalaspGlnProvalHisasp 82  
QY 404 CAAGCACTGATCCCAACTACATGATGATGATGATGATGATGATGATGATGATGAT 463  
D 83 GlnuetaiaProserSerGlnpHeGlnIleSerValalaspGlnProvalHisasp 102  
QY 464 GGAACAACAACATTTCTTCAACGAGTTCCTGCTGATGCTGATGCTGATGCTGATG 523  
D 103 GlnuetaiaProserSerGlnpHeGlnIleSerValalaspGlnProvalHisasp 122  
QY 524 AATCTGTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 577  
D 123 AsnthrValtrpansnGlnalalaserApsCyspHeatrgpHeGlnaspTrpAlaser 140



```

RESULT 7
ABB65017 standard; Protein: 133 AA.
ID ABB65017 standard; Protein: 133 AA.
AC ABB65017;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 21843.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW.
XX
XX WPI: 2001-656860/75.
XX
XX N-PSDB; ABL09120.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 21843; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
XX sequences (AB101840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 133 AA;

Alignment Scores:
Pred. No.: 7.29e-29 Length: 133
Score: 310.00 Matches: 57
Percent Similarity: 60.90% Conservative: 24
Best Local Similarity: 42.86% Mismatches: 48
Query Match: 15.40% Indels: 4
Gaps: 1

US-09-763-902b-15 (1-1146) x ABB65017 (1-133)
OY 170 GATTTAACTTATGTAGATCAGCATGTCGTCGAGATTGTCATATTTAC 229
Db 4 AspLeuYsAlaLysValIGluserCysAlaArgThrAlaAspTrpMetThrArgLeuTyr 23
OY 230 TATGACACATGATTAAGAGAGGSCACTTAACAGGCTGATTCGACAGCCACC 289
Db 24 TyrAlaSerValaAspAsnArgArgGlnGlnIleGlyArgLeuTyrLeuAspAsnAlaThr 43
OY 290 TTAATATGATGATGAATGCGCTTCAGGCTGATGCCCAATAATATTTTGGACACA 349
Db 44 LeuSerTrpAsnGlyAsnGlyAlaIleGlyArgGlnMetIleGlyLeuSerTyrPheGlnGlu 63

```

```

OY 350 TTGCCTTTAGTAGAGTTCAGGTCATATATGTAGATGCCAACAGTTGAGCAAGA 409
Db 64 LeuProSerSerAsnHisGlnLeuAsnThrLeuAspIleProIleValaAspGlnAla 83
OY 410 ACTCAGTCCCAACTACAGTGTGTTGTGACACAGTCGAGACTGTGAAGTTGATGAAAC 469
Db 84 ValSerAsnGlnLeuAlaTyrLeuIleMetAlaSerGlySerValIlyspheAlaAspGln 103
OY 470 AAACACATTTCTTCAACAGCAACTTCCTGCTGACTGCTGACATCCCAACATACT 529
Db 104 GlnLeuArgLysPheGlnGlnThrPheIleValThrAlaGlu-----AsnAsp 119
OY 530 GTGTGAGAGATTGCAAGTGAATGCTTCCTGCTTCCATCAAGAT 568
Db 120 LysTrpLysValValSerAspCysTyrArgMetGlnGlu 132

RESULT 8
ABB67404 standard; Protein: 688 AA.
ID ABB67404 standard; Protein: 688 AA.
AC ABB67404;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 29004.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW.
XX
XX WPI: 2001-656860/75.
XX
XX N-PSDB; ABL11507.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 29004; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
XX sequences (AB101840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 688 AA;

Alignment Scores:
Pred. No.: 0.000184 Length: 688
Score: 114.00 Matches: 43
Percent Similarity: 39.86% Conservative: 16
Best Local Similarity: 29.05% Mismatches: 63

```

```

Query Match: 5.66% IndeIs: 26
DB: 22 Gaps: 6
US-09-763-902B-15 (1-1146) x ABB67404 (1-688)
QY 140 GTCACAGAGGCGCCAGATGGCCAGCTGTCTGGATTTTAAACATTAATGATGATGAGCATGT 199
    |||||:|||||:|:|:
Db 3 AlathrgInserGIcGlnProserProGInser----- 13
QY 200 AGAGCTGCTGAAGACTTTTCTCAATATTTCATATGACACATGATGAATAAGAGACGGCA 259
    |||||:|||||:|||||:|:|:
Db 14 ---ValGIuArvgIuPheValArgGIuInTyTyrThIleLeuAsnIysAlaProAsnHis 32
QY 260 CTAAACAGGCTGTATGTGACACAGGCCACCTTAATA---TGAATGGAATGCTGTTC 316
    ||| ||| ||| :|||: ||| |||
Db 33 LeuHisArgPheTyAsnHisAsnSerSerTyIleHisGIuGIuSerIysIleuVal 52
QY 317 GGGCGGATGGCCCTAAATAATTTTGTGACACATTCGCCCTTAAGTAGATGCCAGTCAT 376
    |||||:|||||:|||||:|:|:
Db 53 GIuGIuArvgIuIleHisAsnArgIleGln-----GlnLeuAsn 65
QY 377 ATGTAGATTCG-----CAACAGCTTCATAGCAACAGCACTAGTCCCAACT 424
    |||||:|||||:|||||:|:|:
Db 66 PheAsnArCyHisAlaIysIleSerGIuValAspAlaGIuAlaThrIleuGIuAsnGIu 85
QY 425 ACAGTCTCTGTGTGACACAGTGGAACTGTGAAGTTGATGGAACAACAACACTTTCTTC 484
    |||||:|||||: ||| :||| |||
Db 86 ValValValGIuValThr---GIuGIuLeuSerAsnAspGIuGIuProMetArgArgPhe 104
QY 485 AACCAAGACTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 544
    |||||:|||||:|||||:|:|:
Db 105 ThrGIuThrPheValLeuAlaAlaGIuInserProIysLys-----TyTyrValHis 121
QY 545 AGTGATGCTTCGCTTTCACAGAT 568
    :|||: |||||:|||||:|:|:
Db 122 AsnAspIlePheArgTyGIuInsp 129
RESULT 9
ABB66052
ID ABB66052 standard; Protein; 690 AA.
AC ABB66052;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 24948.
XX
XX Drosophila developmental biology; cell signalling; insecticide;
XX
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX MO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001MO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX
XX N-PSDB; ABL10153.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX
XX interactions -
XX
XX Disclosure: SEQ ID NO 24948; 21pp + Sequence Listing; English..

```

CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI6176-ABL30511), expressed DNA sequences (ABLO1840-ABLI6175) and the encoded proteins.
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
DB	
SQ	Sequence    690 AA;
Alignment Scores:	
Pred. No.:         0 000185	Length:                  690
Score:              114.00	Matches:                 43
Percent Similarity: 39.86%	Conservative:            16
Best Local Similarity: 29.05%	Mismatches:             63
Query Match:         5.66%	Gaps:                    26
	Indels:                    6
US-09-763-902B-15 (1-1146) x ABB66032 (1-650)	
OY	140 GCTACAGGTCGCCAGATGGCCAGCCTCTGTGAATTAAACTTATGTAGATCAGGCATGT..199      :::              ::  Db      5 AlathrcInserIncInProsePrroginSer.....15
OY	200 AGAAGCTCGTAGCAAGTTTGATTAATTCTATGAGACAATGGATAAAAAGAAGCGGCA..255                                                           Db      16 ---ValcIyArgrgluhPhevalrPglnTrTyrrThLeuLysAlaProAsnHis  34
OY	260 CTAAACAGCGCTGATATGTGACAAAGCGCACCTTAATA---TGGAAATGMAATGCTGTTCA..316                       :         :         :         :         : Db      35 LeuHIsArGrHeTYrSnhISasnsErSeTYrlIHLSgLYlUsErLYleuVaIvaI  54
OY	317 GGCGTGGATCCCTTAATAATTTTTTTGACACATTTCCCTTAGTAGTCCAGATCAT  376 GGG         :         :                                         Db      55 GLyLnArgrgluIrLenHSsnArGlLegln.....GlnLeuasn  67
OY	377 ATGTtAAATtGC-----CACCAAGTTCATGACACAAGCAACTCACTCCCAAAct..428                                                                   Db      68 PheaSnApSpShISalALeLySIleserGIvLaSPalaGLnaIatHrLeugLYAnslY  87
OY	425 ACAGCTCTCTTTGTGACAGCTGSAAGCTGGAAGTTGATGGAAACAACACATTCTTC  488    :::             ::                                Db      88 ValVaIvaIcInValThr---GLyGUlsLeuSerASnASPcIGlnPrometArGARpHe  106
OY	485 AACCAAGAACTTCCTGGCTGAGCTGCTCAAGCTCCCAACAATACTGTGTGAGAGATTGCA..544                                                                 Db      107 ThrGlnThrPheValLeuaIaaIagInserrProlySLys-----TYrrYvaIHIs  122
OY	545 AGTGATTTGCTTCGTTTTCAGAT  568 ::          :::      Db      124 AsnaSpllephearGYrGlnasp  131
RESULT 10	
AAGI2073	ID AAGI2073 standard; Protein; 125 AA.
XX	AAGI2073;
XX	17-Oct-2000 (first entry)
XX	Arabidopsis thaliana protein fragment SEO ID NO: 11042.
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.
XN	Arabidopsis thaliana.
XS	

PN EPI033405-A2.  
 XX 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 XX 25-FEB-1999; 9905-0121825.  
 PR 05-MAR-1999; 9905-0123180.  
 PR 09-MAR-1999; 9905-0123548.  
 PR 23-MAR-1999; 9905-0125788.  
 PR 25-MAR-1999; 9905-0126264.  
 PR 29-MAR-1999; 9905-0126785.  
 PR 01-APR-1999; 9905-0127462.  
 PR 08-APR-1999; 9905-0128234.  
 PR 16-APR-1999; 9905-0129845.  
 PR 19-APR-1999; 9905-0130077.  
 PR 21-APR-1999; 9905-0130449.  
 PR 23-APR-1999; 9905-0130510.  
 PR 23-APR-1999; 9905-0130891.  
 PR 28-APR-1999; 9905-0131449.  
 PR 30-APR-1999; 9905-0132048.  
 PR 30-APR-1999; 9905-0132407.  
 PR 04-MAY-1999; 9905-0132484.  
 PR 05-MAY-1999; 9905-0132485.  
 PR 06-MAY-1999; 9905-0132486.  
 PR 07-MAY-1999; 9905-0132487.  
 PR 11-MAY-1999; 9905-0132863.  
 PR 14-MAY-1999; 9905-0134218.  
 PR 14-MAY-1999; 9905-0134219.  
 PR 14-MAY-1999; 9905-0134221.  
 PR 14-MAY-1999; 9905-0134370.  
 PR 18-MAY-1999; 9905-0134768.  
 PR 19-MAY-1999; 9905-0134941.  
 PR 20-MAY-1999; 9905-0135114.  
 PR 21-MAY-1999; 9905-0135353.  
 PR 24-MAY-1999; 9905-0135629.  
 PR 25-MAY-1999; 9905-0136021.  
 PR 27-MAY-1999; 9905-0136392.  
 PR 28-MAY-1999; 9905-0136782.  
 PR 01-JUN-1999; 9905-0137222.  
 PR 03-JUN-1999; 9905-0137528.  
 PR 04-JUN-1999; 9905-0137502.  
 PR 07-JUN-1999; 9905-0137724.  
 PR 08-JUN-1999; 9905-0138094.  
 PR 10-JUN-1999; 9905-0138547.  
 PR 10-JUN-1999; 9905-0138847.  
 PR 14-JUN-1999; 9905-0139119.  
 PR 16-JUN-1999; 9905-0139452.  
 PR 16-JUN-1999; 9905-0139453.  
 PR 17-JUN-1999; 9905-0139454.  
 PR 18-JUN-1999; 9905-0139454.  
 PR 18-JUN-1999; 9905-0139455.  
 PR 18-JUN-1999; 9905-0139456.  
 PR 18-JUN-1999; 9905-0139457.  
 PR 18-JUN-1999; 9905-0139458.  
 PR 18-JUN-1999; 9905-0139459.  
 PR 18-JUN-1999; 9905-0139460.  
 PR 18-JUN-1999; 9905-0139461.  
 PR 18-JUN-1999; 9905-0139462.  
 PR 18-JUN-1999; 9905-0139463.  
 PR 18-JUN-1999; 9905-0139750.  
 PR 18-JUN-1999; 9905-0139763.  
 PR 21-JUN-1999; 9905-0139817.  
 PR 22-JUN-1999; 9905-0139839.  
 PR 23-JUN-1999; 9905-0140353.  
 PR 23-JUN-1999; 9905-0140354.  
 PR 24-JUN-1999; 9905-0140695.  
 PR 28-JUN-1999; 9905-0140823.  
 PR 29-JUN-1999; 9905-0140991.  
 PR 30-JUN-1999; 9905-0141287.  
 PR 01-JUL-1999; 9905-0141842.

PR 01-JUL-1999; 9905-0142154.  
 PR 02-JUL-1999; 9905-0142055.  
 PR 06-JUL-1999; 9905-0142390.  
 PR 08-JUL-1999; 9905-0142803.  
 PR 09-JUL-1999; 9905-0142920.  
 PR 12-JUL-1999; 9905-0142977.  
 PR 13-JUL-1999; 9905-0143542.  
 PR 14-JUL-1999; 9905-0143624.  
 PR 15-JUL-1999; 9905-0144005.  
 PR 16-JUL-1999; 9905-0144085.  
 PR 16-JUL-1999; 9905-0144086.  
 PR 19-JUL-1999; 9905-0144335.  
 PR 19-JUL-1999; 9905-0144335.  
 PR 19-JUL-1999; 9905-0144331.  
 PR 19-JUL-1999; 9905-0144332.  
 PR 19-JUL-1999; 9905-0144333.  
 PR 19-JUL-1999; 9905-0144334.  
 PR 19-JUL-1999; 9905-0144335.  
 PR 20-JUL-1999; 9905-0144352.  
 PR 20-JUL-1999; 9905-0144632.  
 PR 20-JUL-1999; 9905-0144884.  
 PR 21-JUL-1999; 9905-0144814.  
 PR 21-JUL-1999; 9905-0145086.  
 PR 21-JUL-1999; 9905-0145088.  
 PR 22-JUL-1999; 9905-0145085.  
 PR 22-JUL-1999; 9905-0145087.  
 PR 22-JUL-1999; 9905-0145089.  
 PR 22-JUL-1999; 9905-0145192.  
 PR 23-JUL-1999; 9905-0145145.  
 PR 23-JUL-1999; 9905-0145218.  
 PR 26-JUL-1999; 9905-0145224.  
 PR 27-JUL-1999; 9905-0145276.  
 PR 27-JUL-1999; 9905-0145591.  
 PR 27-JUL-1999; 9905-0145591.  
 PR 28-JUL-1999; 9905-0145951.  
 PR 02-AUG-1999; 9905-0146386.  
 PR 02-AUG-1999; 9905-0146388.  
 PR 02-AUG-1999; 9905-0146389.  
 PR 03-AUG-1999; 9905-0147038.  
 PR 04-AUG-1999; 9905-0147204.  
 PR 04-AUG-1999; 9905-0147302.  
 PR 05-AUG-1999; 9905-0147192.  
 PR 05-AUG-1999; 9905-0147260.  
 PR 06-AUG-1999; 9905-0147303.  
 PR 06-AUG-1999; 9905-0147416.  
 PR 09-AUG-1999; 9905-0147483.  
 PR 09-AUG-1999; 9905-0147935.  
 PR 10-AUG-1999; 9905-0148171.  
 PR 11-AUG-1999; 9905-0148319.  
 PR 12-AUG-1999; 9905-0148341.  
 PR 13-AUG-1999; 9905-0148565.  
 PR 13-AUG-1999; 9905-0148684.  
 PR 16-AUG-1999; 9905-0149388.  
 PR 17-AUG-1999; 9905-0149175.  
 PR 18-AUG-1999; 9905-0149426.  
 PR 20-AUG-1999; 9905-0149722.  
 PR 20-AUG-1999; 9905-0149723.  
 PR 20-AUG-1999; 9905-0149929.  
 PR 23-AUG-1999; 9905-0149902.  
 PR 23-AUG-1999; 9905-0149930.  
 PR 25-AUG-1999; 9905-0150566.  
 PR 26-AUG-1999; 9905-0150884.  
 PR 27-AUG-1999; 9905-0151065.  
 PR 27-AUG-1999; 9905-0151066.  
 PR 27-AUG-1999; 9905-0151080.  
 PR 30-AUG-1999; 9905-0151303.  
 PR 31-AUG-1999; 9905-0151308.  
 PR 01-SEP-1999; 9905-0151930.  
 PR 07-SEP-1999; 9905-0152363.  
 PR 10-SEP-1999; 9905-0153070.  
 PR 13-SEP-1999; 9905-0153758.  
 PR 15-SEP-1999; 9905-0154018.  
 PR 16-SEP-1999; 9905-0154039.

PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157573.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161921.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:  
 Pred. No.: 0.0001  
 Score: 113.50  
 Percent Similarity: 46.02%  
 Best Local Similarity: 27.43%  
 Query Match: 5.64%  
 DB: 21  
 Gaps: 3

US-09-763-902b-15 (1-1146) x AMG12073 (1-125)  
 QY 206 GCTGAGAGAGTTGTCATATTTACTATGAGACAATGATAAAGAGAGGACTAAC 265  
 Db 11 AAlaSerAlaPheValAsnHisTyrTyrHisLeuPheAspAsnAspArgSerSerLeuSer 30  
 QY 266 AGGCTGATCTGAGACAGGCGCCCTTAATATGATGAGAAATGCTGTTCAGGGCTGAT 325  
 Db 31 SerLeuTyrAsnProHisLeuSerLeuThrPheGluGluGlnThrHisLeuTyrGlyValAsp 50  
 QY 326 GCGCTAAATATTTTGTGACACATGCGCTTCTAGTGAAGTTCACG-----GTCAATATG 379  
 Db 51 AsnIleSerAlaHisLeuTyrGlnLeuProPheAspGlnCysHisHisLeuIleSerThr 70  
 QY 380 TTAGATGCCACAGAGTTGATGAGACAGACAGCTGAGTCCAAAGTCTTTGTG 439  
 Db 71 ValAspSerGlnPro-----SerSerMetAlaGlyGlyCysGlyGlyLeuValPhe 88  
 QY 440 ACCAGTGGAGACTGGAAGTTGAGAGAAACAACAATTC-----TTCACAGAGACTTC 496  
 Db 89 ValSerGlySerIleGlnLeuHisGlyGlnAspHisProLeuAspArgPheSerGlnValTyr 108  
 QY 497 CTGCTAGCTGCTAGTGCACATGCCAACACATCTGTGTG 535

Db 109 LeuLeuSerIleCysAsnArgPheGluGlyPheIleTyr 121  
 RESULT 11  
 ID AAG05145 standard; Protein; 153 AA.  
 AC AAG05145;  
 DT 17-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 1441.  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 OS Arabidopsis thaliana.  
 PN EP1033405-A2.  
 PD 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF 25-FEB-1999; 99US-0121825.  
 XX 05-MAR-1999; 99US-0122180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135621.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139751.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144844.  
PR 21-JUL-1999; 99US-0144844.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145921.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148564.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150366.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160960.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:  
Pred. No.: 0.000109 Length: 153  
Score: 113.50 Matches: 37  
Percent Similarity: 45.32% Conserved: 26  
Best Local Similarity: 25.908 Mismatches: 65  
Query Match: 5.64% Indels: 11  
DB: 21 Gaps: 5

US-09-763-902b-15 (1-1146) x ANG05145 (1-153)

OY 158 GCCACGCTCTGATTTAACTTATGTAGATCAGCAGTGTAGAGCTGCTGAGGACTTT 217  
Db 13 SerAlaserleuserleuysMetSerGlnMetAspProAspAlaValSerIysAlaPhe 32

PR	18-MAY-1999	9905-0134768
PR	19-MAY-1999	9905-0135141
PR	20-MAY-1999	9905-0135194
PR	21-MAY-1999	9905-0135353
PR	24-MAY-1999	9905-0135659
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136367
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137228
PR	03-JUN-1999	9905-0137558
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138340
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139411
PR	16-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139442
PR	18-JUN-1999	9905-0139447
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139458
PR	18-JUN-1999	9905-0139459
PR	18-JUN-1999	9905-0139460
PR	18-JUN-1999	9905-0139461
PR	18-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139464
PR	18-JUN-1999	9905-0139465
PR	18-JUN-1999	9905-0139466
PR	18-JUN-1999	9905-0139467
PR	18-JUN-1999	9905-0139468
PR	18-JUN-1999	9905-0139469
PR	18-JUN-1999	9905-0139470
PR	18-JUN-1999	9905-0139471
PR	18-JUN-1999	9905-0139472
PR	18-JUN-1999	9905-0139473
PR	18-JUN-1999	9905-0139474
PR	18-JUN-1999	9905-0139475
PR	18-JUN-1999	9905-0139476
PR	18-JUN-1999	9905-0139477
PR	18-JUN-1999	9905-0139478
PR	18-JUN-1999	9905-0139479
PR	18-JUN-1999	9905-0139480
PR	18-JUN-1999	9905-0139481
PR	18-JUN-1999	9905-0139482
PR	18-JUN-1999	9905-0139483
PR	18-JUN-1999	9905-0139484
PR	18-JUN-1999	9905-0139485
PR	18-JUN-1999	9905-0139486
PR	18-JUN-1999	9905-0139487
PR	18-JUN-1999	9905-0139488
PR	18-JUN-1999	9905-0139489
PR	18-JUN-1999	9905-0139490
PR	18-JUN-1999	9905-0139491
PR	18-JUN-1999	9905-0139492
PR	18-JUN-1999	9905-0139493
PR	18-JUN-1999	9905-0139494
PR	18-JUN-1999	9905-0139495
PR	18-JUN-1999	9905-0139496
PR	18-JUN-1999	9905-0139497
PR	18-JUN-1999	9905-0139498
PR	18-JUN-1999	9905-0139499
PR	18-JUN-1999	9905-0139500
PR	18-JUN-1999	9905-0139501
PR	18-JUN-1999	9905-0139502
PR	18-JUN-1999	9905-0139503
PR	18-JUN-1999	9905-0139504
PR	18-JUN-1999	9905-0139505
PR	18-JUN-1999	9905-0139506
PR	18-JUN-1999	9905-0139507
PR	18-JUN-1999	9905-0139508
PR	18-JUN-1999	9905-0139509
PR	18-JUN-1999	9905-0139510
PR	18-JUN-1999	9905-0139511
PR	18-JUN-1999	9905-0139512
PR	18-JUN-1999	9905-0139513
PR	18-JUN-1999	9905-0139514
PR	18-JUN-1999	9905-0139515
PR	18-JUN-1999	9905-0139516
PR	18-JUN-1999	9905-0139517
PR	18-JUN-1999	9905-0139518
PR	18-JUN-1999	9905-0139519
PR	18-JUN-1999	9905-0139520
PR	18-JUN-1999	9905-0139521
PR	18-JUN-1999	9905-0139522
PR	18-JUN-1999	9905-0139523
PR	18-JUN-1999	9905-0139524
PR	18-JUN-1999	9905-0139525
PR	18-JUN-1999	9905-0139526
PR	18-JUN-1999	9905-0139527
PR	18-JUN-1999	9905-0139528
PR	18-JUN-1999	9905-0139529
PR	18-JUN-1999	9905-0139530
PR	18-JUN-1999	9905-0139531
PR	18-JUN-1	

```
PR 27-JUL-1999: 99US-0145919.
PR 28-JUL-1999: 99US-0145951.
PR 02-AUG-1999: 99US-0146386.
PR 02-AUG-1999: 99US-0146388.
PR 02-AUG-1999: 99US-0146389.
PR 03-AUG-1999: 99US-0147203.
PR 04-AUG-1999: 99US-0147204.
PR 04-AUG-1999: 99US-0147302.
PR 05-AUG-1999: 99US-0147192.
PR 05-AUG-1999: 99US-0147260.
PR 06-AUG-1999: 99US-0147303.
PR 06-AUG-1999: 99US-0147416.
PR 09-AUG-1999: 99US-0147493.
PR 09-AUG-1999: 99US-0147933.
PR 10-AUG-1999: 99US-0148171.
PR 11-AUG-1999: 99US-0148319.
PR 12-AUG-1999: 99US-0148341.
PR 13-AUG-1999: 99US-0148565.
PR 13-AUG-1999: 99US-0148684.
PR 16-AUG-1999: 99US-0149368.
PR 17-AUG-1999: 99US-0149175.
PR 18-AUG-1999: 99US-0149426.
PR 20-AUG-1999: 99US-0149722.
PR 20-AUG-1999: 99US-0149723.
PR 20-AUG-1999: 99US-0149929.
PR 23-AUG-1999: 99US-0149902.
PR 23-AUG-1999: 99US-0149930.
PR 25-AUG-1999: 99US-0150566.
PR 26-AUG-1999: 99US-0150884.
PR 27-AUG-1999: 99US-0151065.
PR 27-AUG-1999: 99US-0151066.
PR 27-AUG-1999: 99US-0151080.
PR 30-AUG-1999: 99US-0151303.
PR 31-AUG-1999: 99US-0151308.
PR 01-SEP-1999: 99US-0151930.
PR 07-SEP-1999: 99US-0152363.
PR 10-SEP-1999: 99US-0153070.
PR 13-SEP-1999: 99US-0153758.
PR 15-SEP-1999: 99US-0154018.
PR 16-SEP-1999: 99US-0154039.
PR 20-SEP-1999: 99US-0154779.
PR 22-SEP-1999: 99US-0155139.
PR 23-SEP-1999: 99US-0155486.
PR 24-SEP-1999: 99US-0155659.
PR 28-SEP-1999: 99US-0156458.
PR 29-SEP-1999: 99US-0156596.
PR 04-OCT-1999: 99US-0157117.
PR 05-OCT-1999: 99US-0157753.
PR 06-OCT-1999: 99US-0157865.
PR 07-OCT-1999: 99US-0158029.
PR 08-OCT-1999: 99US-0158232.
PR 12-OCT-1999: 99US-0158369.
PR 13-OCT-1999: 99US-0159293.
PR 13-OCT-1999: 99US-0159294.
PR 13-OCT-1999: 99US-0159295.
PR 14-OCT-1999: 99US-0159329.
PR 14-OCT-1999: 99US-0159330.
PR 14-OCT-1999: 99US-0159331.
PR 14-OCT-1999: 99US-0159637.
PR 14-OCT-1999: 99US-0159638.
PR 18-OCT-1999: 99US-0159584.
PR 18-OCT-1999: 99US-0159584.
PR 21-OCT-1999: 99US-0160741.
PR 21-OCT-1999: 99US-0160767.
PR 21-OCT-1999: 99US-0160768.
PR 21-OCT-1999: 99US-0160770.
PR 21-OCT-1999: 99US-0160814.
PR 21-OCT-1999: 99US-0160815.
PR 22-OCT-1999: 99US-0160980.
PR 22-OCT-1999: 99US-0160981.
PR 23-OCT-1999: 99US-0160989.
PR 25-OCT-1999: 99US-0161404.
PR 25-OCT-1999: 99US-0161405.
PR 25-OCT-1999: 99US-0161406.

PR 26-OCT-1999: 99US-0161359.
PR 26-OCT-1999: 99US-0161360.
PR 26-OCT-1999: 99US-0161361.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161920.
PR 28-OCT-1999: 99US-0161920.
PR 29-OCT-1999: 99US-0162142.

Alignment Scores:
Pred. No.: 0.000132
Score: 112.50
Percent Similarity: 47.93%
Best Local Similarity: 27.27%
Query Match: 5.59%
DB: 21
Gaps: 5

US-09-763-902B-15 (1-1146) x AAC08682 (1-123)

QY 206 GCTGAGAGTTGTCAATTTACTTGTGACAAATGATTAAGAGAGGCGCACTAAC 265
      ::::: ||||| ||||| ||||| ||||| |||||
DB 7 SerLysAlaPheValGluHisTyrSerThrPheAspThrAsnArgValGlyLeuAla 26
QY 266 AGCGTGTATCTGACAGAGCCCACTTAATGATGAATGAATGCTTTACAGGCTGAT 325
      ||||| ::::: ||||| ||||| ||||| |||||
DB 27 GlyLeuTyrGlnGluAlaSerMetLeuThrPheGluGlyGlnLysIleGlnValGln 46
QY 326 GCCCTAATAATATTTTGTGACACATTGCCT-----TTAGTGTAGTTCCAGTCAATATG 379
      ::::: ||||| ||||| ||||| ||||| |||||
DB 47 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnLysLysHisIleSerThr 66
QY 380 TTAGATTGCCAACCACCTTATATGACAGACACATCCCAACTCACTTCTGTGTGG 439
      ::||| ||||| ||||| ||||| ||||| |||||
DB 67 ValAspCysGlnPro-----SerGlyProAlaSerLysLeuValPhe 81
QY 440 ACCAGTGAAGTGTGAGTTTGTGAGAACAAACAACATTTTC--TTCAACCGAAGCTTC 496
      ||||| ::::: ||||| ||||| ||||| |||||
DB 82 ValSerGlyAsnLeuGlnIleuAlaGlyGlnGlnHisAlaLeuLysPheSerGlnMetPhe 101
QY 497 CTGCTGACCTGCTGACGACCTCCCAACATTAATGATGAGATGATGATGATGCTTC 556
      ||||| ||||| ||||| ||||| ||||| |||||
DB 102 HisLeu-----MetProThrProGlnGlySerPheTyr--ValPheAsnAspIlePhe 118
QY 557 CGT 559
      |||
DB 119 Arg 119

RESULT 13
AAC08681
ID AAC08681 standard; Protein; 126 AA.
XX
AC AAC08681:
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6316.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay: genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
OS
PN EP1033405-A2.
PN
PD 06-SEP-2000.
XX
PE 25-FEB-2000: 2000EP-0301439.
XX
PF 25-FEB-1999: 99US-0121825.
XX
PR 05-MAR-1999: 99US-0123180.
PR 09-MAR-1999: 99US-0123548.
PR 23-MAR-1999: 99US-0125788.
PR 25-MAR-1999: 99US-0126264.
```

PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140655.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147483.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.





PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0145386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 0.000136  
Score: 112.50  
Percent Similarity: 47.97%  
Best Local Similarity: 26.83%  
Query Match: 5.59%  
DB: 21  
Gaps: 5  
Indels: 11  
Matches: 33  
Conservative: 26  
Mismatch: 53  
Indels: 11

US-09-763-902b-15 (1-1146) x AAG05147 (1-131)

OY 206 GCTGAGAGCTTGTCAATATTCTATGAGCAATGATGATAAAGAGCGGCGCTACACC 265  
DB 7 SerLysAlaPheValGlnHisItyrSerThrPheAspThrAsnArgValGlnIleuAla 26  
OY 266 AGGCTGATCTGAGCAAGCCACCTTAATATGATGAAATGCAATGCTTTACAGCGCTGAT 325  
DB 27 GlyLeuTygGlnIleAlaSerMetLeuThrPheGlnGlnIleGlnIleGlnIleGln 46  
OY 326 GCCCTAATAATATTTTTCACACATTCCTCTCTAGAGTTCACAGTCAATATG 379  
DB 47 SerIleValAlaIleLeuThrSerLeuProPheGlnGlnIleGlnIleGlnIleSerThr 66

```
QY 380 TTGATGGCCACCACTTCATGACGACCACTCACTCCCAACTGATCTTGTGTG 439
Db 67 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 81
QY 440 ACCAGTGGAACTGTGATGTTGATGAAACAACAACATTC--TTCAACGAGAACTTC 496
Db 82 ValSerGlyAsnLeuGlnLeuAlaGlyGlyValHisAlaLeuLysPheSerGlnMetPhe 101
QY 497 CTCCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGCTC 556
Db 102 HisLeu-----MetProHisProGlnGlySerPheTyr--ValPheHisAsnAlaPhe 118
QY 557 CGTTTCAA 565
Db 119 SerTrpArg 121

RESULT 15
AG05146
ID AG05146 standard; Protein; 134 AA.
XX
AC AG05146;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1442.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145277.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
```

PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 06-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151338.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154329.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160880.  
 PR 22-OCT-1999; 99US-0160881.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:  
 Pred. No.: 0.000137 Length: 134  
 Score: 112.50 Matches: 33  
 Percent Similarity: 47.978 Conservative: 26  
 Best Local Similarity: 26.838 Mismatches: 53  
 Query Match: 5.598 Indels: 11  
 DB: 21 Gaps: 5

US-09-763-902b-15 (1-1146) x AAG05146 (1-134)

QY 206 GCTGAGAGCTTGTCTCAATATTACATGAGCAATGATMAAAGAGAGCGGCACTAAC 265  
 : : : : :  
 Db 10 SerLysAlaPheValGlnHisTyrTyrSerThrPheAspThrAsnArgValGlyLeuAla 29  
 : : : : :  
 QY 266 AGCGTGTATCTGGACAGGCGCACTTAATATGGAATGAAATGCTGTTCCAGGGCTGAT 325  
 : : : : :  
 Db 30 GlyLeuTyrGlnGlnLysSerMetLeuThrPheGlnGlyGlnLysIleGlnGlyValGln 49  
 : : : : :  
 QY 326 GCCCTAATAATTTTGTACACATTCCT-----TCTAGTGAAGTCCAGGTCATATG 379  
 : : : : :  
 Db 50 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnCysLysHisIleSerThr 69  
 : : : : :  
 QY 380 TTAGATTGCCAACCACTTCATGAGCAAGCACTGACCTCCAACTACAGTCTTGTGTG 439  
 : : : : :  
 Db 70 ValAspCysGlnPro-----SerGlyProHisSerGlyMetLeuValPhe 84  
 : : : : :  
 QY 440 ACCAGTGAAGCTGTGATGATGGAACAACAACATTC--TTCACCAAGAACTTC 496  
 : : : : :  
 Db 85 ValSerGlyAsnLeuGlnLeuAlaGlyGlnLysHisAlaLeuLysSerGlnMetPhe 104  
 : : : : :  
 QY 497 CTCCTGACGCTCAGTCCAGTCCCAACATACGCTGTGGAAGTTCAGATTCCTTC 556  
 : : : : :  
 Db 105 HisLeu-----MetProThrProGlnGlySerPheTyr--ValPheAsnAspIlePhe 121  
 : : : : :  
 QY 557 CGTTTCAA 565  
 : : : : :  
 Db 122 SerTirParg 124

Search completed: June 24, 2003, 19:31:07  
 Job time : 58.8062 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 01:01:07 ; Search time 138.193 Seconds  
(Without alignments) 8915.714 Million cell updates/sec

Title: US-09-763-902B-10

Perfect score: 830  
Sequence: 1 tgactttggtacacgtg.....ataattataataacacgtg 830

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

## Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCOT\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCOTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	761	91.7	864	10	US-09-764-864-394	Sequence 394, App	
2	318	38.3	318	9	US-09-796-692-4160	Sequence 4160, App	
3	318	38.3	318	9	US-10-040-862-4160	Sequence 4160, App	
4	317	38.2	318	10	US-09-815-343-1201	Sequence 1201, App	
5	250	30.1	1387	10	US-09-764-864-393	Sequence 393, App	
6	216.8	26.1	415	9	US-09-918-995-2901	Sequence 2901, App	
7	171.8	20.7	417	9	US-09-918-995-6557	Sequence 6557, App	
8	113	13.6	8918	10	US-09-764-864-1682	Sequence 1682, App	
9	113	13.6	8919	10	US-09-764-864-1683	Sequence 1683, App	
C	8	113	13.6	8919	10	US-09-918-995-36691	Sequence 36691, App
C	10	36.4	4.4	415	9	US-09-799-999-3	Sequence 3, App1
11	36.4	4.4	88191	10	US-09-764-877-3374	Sequence 3374, App	
12	35.8	4.3	32191	10	US-09-764-877-3374	Sequence 3374, App	
C	13	34.8	4.2	1141	9	US-10-008-118A-11	Sequence 11, App
C	14	34.8	4.2	1141	9	US-09-443-704-11	Sequence 11, App
C	15	34.8	4.2	1490	9	US-10-097-065-31	Sequence 31, App1
16	34.8	4.2	172637	10	US-09-805-458A-3	Sequence 3, App1	
C	17	34.2	4.1	555	9	US-10-123-155-72	Sequence 72, App1
C	18	34	4.1	499	9	US-09-918-995-19696	Sequence 19696, App
19	33.6	4.0	1753	9	US-09-986-480-149	Sequence 149, App	

20	33.6	4.0	5512	9 US-10-239-676-150	Sequence 150, App
21	33.6	4.0	684973	10 US-09-263-959-1	Sequence 1, App11
22	33.2	4.0	1294	9 US-09-938-842A-4254	Sequence 4254, App
23	33.2	4.0	1907	9 US-09-924-340-53	Sequence 53, App1
24	33.2	4.0	1907	9 US-09-992-600A-53	Sequence 53, App1
25	33.2	4.0	1907	9 US-10-000-489-53	Sequence 53, App1
26	33.2	4.0	1907	9 US-10-000-986-53	Sequence 53, App1
27	33.2	4.0	1581	9 US-10-172-086-81	Sequence 81, App1
28	33.2	4.0	482	10 US-09-864-761-15713	Sequence 15713, App
29	33	4.0	4059	12 US-10-042-417-47	Sequence 47, App1
30	33	4.0	4221	9 US-10-125-237-70	Sequence 70, App1
31	33	4.0	4221	9 US-10-105-891-70	Sequence 70, App1
32	33	4.0	8764	9 US-09-764-891-10019	Sequence 10019, App
33	33	4.0	8911	12 US-10-044-090-17	Sequence 17, App1
34	32.8	4.0	1503841	9 US-09-946-807-1	Sequence 1, App11
35	32.8	4.0	1503841	10 US-09-795-668-1	Sequence 1, App11
36	32.8	4.0	1503841	10 US-09-795-668-1	Sequence 1, App11
37	32.4	3.9	216	10 US-09-960-352-7902	Sequence 7902, App
38	32.2	3.9	239	10 US-09-919-580-208	Sequence 208, App
39	32.2	3.9	250	10 US-09-864-761-31695	Sequence 31695, App
40	32	3.9	457	9 US-09-918-995-11773	Sequence 11773, App
41	32	3.9	504	12 US-10-001-843-14	Sequence 14, App1
42	32	3.9	517	10 US-09-919-580-874	Sequence 874, App
43	32	3.9	756	9 US-10-123-155-392	Sequence 392, App
44	32	3.9	1194	10 US-09-834-975-840	Sequence 840, App
45	32	3.9	3285	9 US-09-986-480-29	Sequence 29, App1

## ALIGNMENTS

RESULT 1  
US-09-764-864-394  
Sequence 394, Application US/09764864  
Patent No. US20020137253A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764,864  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 394  
LENGTH: 864  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (7)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (9)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (15)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (31)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (837)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (653)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (863)  
OTHER INFORMATION: n equals a,t,g, or c

US-09-764-864-394

Query Match 91.7% Score 761; DB 10; Length 864;  
Best Local Similarity 100.0%; Pred. No. 4,9e-240;  
Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

70 GGAATACCCCTGGAGAGCCCTCTTCATGAAACACAGATGGCATCTGTGATTCA 129  
|||||  
35 GGAATACCCCTGGAGAGCCCTCTTCATGAAACACAGATGGCATCTGTGATTCA 94  
130 GACCTATGTGATAGAGGCTGACAGGCTGAGAGAGTTTCTCATGCTCTACACAC 189  
|||||  
95 GACCTATGTGATAGAGGCTGACAGGCTGAGAGAGTTTCTCATGCTCTACACAC 154  
190 CATGATTAAGGCGGGGCTTGTGCTCCGCTGTACATGGGACAGCACCGCTGTG 249  
155 CATGATTAAGGCGGGGCTTGTGCTCCGCTGTACATGGGACAGCACCGCTGTG 214  
250 GAATGCAATGCTGTTTCAGACAAAGATCCTTGAAGTATTTTGAATGTGCTTC 309  
|||||  
215 GAATGCAATGCTGTTTCAGACAAAGATCCTTGAAGTATTTTGAATGTGCTTC 274  
310 CAGGAGTTCATCAATCAGCGGTAGACTGACACCTGTTTATGAGCCACACAG 369  
275 CAGGAGTTCATCAATCAGCGGTAGACTGACACCTGTTTATGAGCCACACAG 334  
370 CAGGAGTTCATCAATCAGCGGTAGACTGACACCTGTTTATGAGCCACACAG 429  
|||||  
335 CAGGAGTTCATCAATCAGCGGTAGACTGACACCTGTTTATGAGCCACACAG 394  
430 GGAATACCCCTGGAGAGCCCTCTTCATGAAACACAGATGGCATCTGTGATTCA 489  
|||||  
395 GGAATACCCCTGGAGAGCCCTCTTCATGAAACACAGATGGCATCTGTGATTCA 454  
490 GATGCAAGTACGCTGCTCCGCTTCCAGAGCTGGGCGGAGTGGGAGAGTGC 549  
455 GATGCAAGTACGCTGCTCCGCTTCCAGAGCTGGGCGGAGTGGGAGAGTGC 514  
550 TCTTGTCTTATTCAGCCCTAGCTCTGAGAAATCAAACTCGACTCTCAAGATGT 609  
515 TCTTGTCTTATTCAGCCCTAGCTCTGAGAAATCAAACTCGACTCTCAAGATGT 574  
610 GAGGACACAAATGTTTCTGTTGTTGGGAGACACTGACAGACTCCACTGTGCGAGGT 669  
575 GAGGACACAAATGTTTCTGTTGTTGGGAGACACTGACAGACTCCACTGTGCGAGGT 634  
670 TGAACCTTTTGTGCTCAGAGTCTAGAGAGTCCCTTCTGTAATATATCTTTGT 729  
635 TGAACCTTTTGTGCTCAGAGTCTAGAGAGTCCCTTCTGTAATATATCTTTGT 694  
730 CATAGTTTCTTTCAGAGTAACTTCTTATTTTCTACTTGGCCAGTGAAGACTC 789  
695 CATAGTTTCTTTCAGAGTAACTTCTTATTTTCTACTTGGCCAGTGAAGACTC 754  
790 TGAATTCGAAATTCGACAAATTAATTAATACATG 830  
755 TGAATTCGAAATTCGACAAATTAATTAATACATG 795

RESULT 2  
US-09-766-692-4160

Sequence 4160, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
CURRENT APPLICATION NUMBER: US/09/796,692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126

PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22,950  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4160  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-4160

Query Match 38.3% Score 318; DB 9; Length 318;  
Best Local Similarity 100.0%; Pred. No. 2.8e-94;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

147 CTTGACAGCTGCTGAGAGTTTGTCAATGCTACTACACCATGATGAGGCGGC 206  
1 CTTGACAGCTGCTGAGAGTTTGTCAATGCTACTACACCATGATGAGGCGGC 60  
207 GTTGTGCTGCTCCGCTTATGATGGGACAGCCCTGCTGTGAAATGCTGTTT 266  
61 GTTGTGCTGCTCCGCTTATGATGGGACAGCCCTGCTGTGAAATGCTGTTT 120  
267 CAGGACAGATCCTTGAAGTATTTTGAATGTTGCTTCCAGGAGTCCAAATCA 326  
121 CAGGACAGATCCTTGAAGTATTTTGAATGTTGCTTCCAGGAGTCCAAATCA 180  
327 GCGTGTAGACTGCGACCTGTTTCATGATGAGGACACACAGGAGTCCCTTG 386  
181 GCGTGTAGACTGCGACCTGTTTCATGATGAGGACACACAGGAGTCCCTTG 240  
387 TTGTGATCTGTGATCAGTAACTTTGAGGGGAAACAAACAGGACTTCAACCAACT 446  
241 TTGTGATCTGTGATCAGTAACTTTGAGGGGAAACAAACAGGACTTCAACCAACT 300  
447 TCAATCTGACCGCCGAG 464  
301 TCAATCTGACCGCCGAG 318

RESULT 3  
US-10-040-862-4160

Sequence 4160, Application US/10040862  
Publication No. US20030078396A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
APPLICANT: Retter, Marc  
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther  
FILE OF INVENTION: Hematological Malignancies  
FILE REFERENCE: 014058-013520US

CURRENT APPLICATION NUMBER: US/10/040,862  
CURRENT FILING DATE: 2001-11-06  
PRIOR APPLICATION NUMBER: US 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: US 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: US 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: US 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: US 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: US 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: US 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: US 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: US 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: US 60/223,378  
PRIOR FILING DATE: 2000-08-07  
PRIOR APPLICATION NUMBER: US 09/796,692  
PRIOR FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 10467  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 4160  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-040-862-4160

Query Match  
Best Local Similarity 38.3%; Score 318; DB 9; Length 318;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 CCTGACAGCTGCTGAGAGTTTGTCAATGCTCTACTACACCACTGATTAAGCGCGC 206  
DB 1 CCTGACAGCTGCTGAGAGTTTGTCAATGCTCTACTACACCACTGATTAAGCGCGC 60  
QY 207 GTTGTCTGCTCCCGCTGTACATGGGACAGCCACCCCTGCTGCAATGCTGTTT 266  
DB 61 GTTGTCTGCTCCCGCTGTACATGGGACAGCCACCCCTGCTGCAATGCTGTTT 120  
QY 267 CAGACAGAATCTCTGTGATGATTTTGAATGTTGCCCTTCCAGCGAGTTCCAAATCA 326  
DB 121 CAGACAGAATCTCTGTGATGATTTTGAATGTTGCCCTTCCAGCGAGTTCCAAATCA 180  
QY 327 GCGGTGTGATGCTGAGCGCTTCTCATGTGTAAGCCACCAAGCGAGTCTCTTG 386  
DB 181 GCGGTGTGATGCTGAGCGCTTCTCATGTGTAAGCCACCAAGCGAGTCTCTTG 240  
QY 387 TTGTCTATCTGTGATGATGATTTTGAAGGGGAAACAAACAGGACTTCAACGAACT 446  
DB 241 TTGTCTATCTGTGATGATGATTTTGAAGGGGAAACAAACAGGACTTCAACGAACT 300  
QY 447 TCATCTGTACCGCCAGG 464  
DB 301 TCATCTGTACCGCCAGG 318

RESULT 4  
US-09-815-343-1201  
Sequence 1201, Application US/09815343  
Patent No. US2001005596A1  
GENERAL INFORMATION:  
APPLICANT: Meagher, Madeline  
APPLICANT: Xu, Jianshun  
APPLICANT: King, Gordon E.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER  
FILE REFERENCE: 210121.504  
CURRENT APPLICATION NUMBER: US/09/815,343  
CURRENT FILING DATE: 2001-03-22  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1201  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: (1)...(318)  
OTHER INFORMATION: n = A,T,C or G  
US-09-815-343-1201

Query Match  
Best Local Similarity 38.2%; Score 317; DB 10; Length 318;  
Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 147 CCTGACAGCTGCTGAGAGTTTGTCAATGCTCTACTACACCACTGATTAAGCGCGC 206  
DB 1 CCTGACAGCTGCTGAGAGTTTGTCAATGCTCTACTACACCACTGATTAAGCGCGC 60  
QY 207 GTTGTCTGCTCCCGCTGTACATGGGACAGCCACCCCTGCTGCAATGCTGTTT 266  
DB 61 GTTGTCTGCTCCCGCTGTACATGGGACAGCCACCCCTGCTGCAATGCTGTTT 120  
QY 267 CAGACAGAATCTCTGTGATGATTTTGAATGTTGCCCTTCCAGCGAGTTCCAAATCA 326  
DB 121 CAGACAGAATCTCTGTGATGATTTTGAATGTTGCCCTTCCAGCGAGTTCCAAATCA 180  
QY 327 GCGGTGTGATGCTGAGCGCTTCTCATGTGTAAGCCACCAAGCGAGTCTCTTG 386  
DB 181 GCGGTGTGATGCTGAGCGCTTCTCATGTGTAAGCCACCAAGCGAGTCTCTTG 240  
QY 387 TTGTCTATCTGTGATGATGATTTTGAAGGGGAAACAAACAGGACTTCAACGAACT 446  
DB 241 TTGTCTATCTGTGATGATGATTTTGAAGGGGAAACAAACAGGACTTCAACGAACT 300  
QY 447 TCATCTGTACCGCCAGG 464  
DB 301 TCATCTGTACCGCCAGG 318

RESULT 5  
US-09-764-864-393  
Sequence 393, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764,864  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 393  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1379)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (1380)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-864-393

Query Match 30.1%; Score 250; DB 10; Length 1387;

Best Local Similarity 71.0%; Pred. No. 1.9e-71;  
Matches 331; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
OY 98 ATAGACAGAGATGGCATCTGTGATTTCAAGACCTGTGTGATCAGGCTGCAGACT 157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 52 ATAGGAGAGGCGGGGATGTGGGATTTTAAACTTATGTAGATCAGCATGTAGACT 111
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 158 GCTGAGAGATTGTTCATATGCTACTACACACAGCATGGATTAAGCGGCGCTTGTCTCC 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 112 GCTGAGAGATTGTTCATATGCTACTACACACATGGATTAAGCGGCGCTTGTCTCC 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 218 GCGCTGTACATGGCCACAGCCCTGTGTGATGTGATGCTGTTCAGACAAGAA 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 172 AGGCTGTATCTGGACAAGCCACTTATATGATGAAATGCTGTTCAGGCTGAT 231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 278 TCCTGTAGTGTGTTTGAATTTGCTTCCAGCAGATTCATACAGCTGTAGAC 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 232 GCGCTAAATATTTTGTGACACATTCCTTGTAGTGTGCTCAGTCAATATGTAGAT 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 338 TGCCAGCCTGTTCATGATGAGCCACACAGCCAGCCAGCCCTTGTGTCTGTCTGT 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 292 TGCCAGCCTGTTCATGATGAGCCACACAGCCAGCCAGCCCTTGTGTCTGTCTGT 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 398 GGATCACTGAAGTTTGAAGGGGAACAACAAGGAGCTTCAACAGACTTCACTGAC 457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 GGAAGTGTGAAGTTTGAAGGGGAACAACAAGGAGCTTCAACAGACTTCACTGAC 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 458 GCCCAGGCTTCAACCCAGCAACAGTGTGGAAGATGCAAGTGTCTTCCCTTCCAG 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 412 GCTGAGTCCACTCCCAACAATGCTGTGTGAGATTCAGATGCTTCCCTTCCAG 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 518 GACTGGCCAGCTAGTGGGGGGGAGAGCTCTTGTCTGCTATTC 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 472 GATGTGCTGTGTGTGTAAAGGGCAAAAGTCCATCTCATTTGCTG 517
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 6  
US-09-918-995-2901  
Sequence 2901, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918, 995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235, 076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2901  
LENGTH: 415  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-918-995-2901

Query Match  
Best Local Similarity 26.1%; Score 216.8; DB 9; Length 415;  
Matches 218; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
OY 41 CCTGACAGCCCTGTGTTCCCAAGCAGAGAAATACCTGTGAGCCCTCTTCATA 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 196 CCGGCGACCCCTGTGTTCCCAAGCAGAGAAATACCTGTGAGCCCTCTTCATA 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 101 GAACAGAGATGATGCTGTGATTTCAAGACCTATGTGATCAGGCTGCAGAGCTGCT 160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 256 GAACAGAGATGATGCTGTGATTTCAAGACCTATGTGATCAGGCTGCAGAGCTGCT 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 161 GAGGAGTTTGTCAATGTCTACTACACACCATGTGTAAGCGGCGCTTGTCTGCCCCG 220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 316 GAGGAGTTTGTCAATGTCTACTACACACCATGTGTAAGCGGCGCTTGTCTGCCCCG 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
OY 221 CTGTACATGGCCACAGCCACCCCTGTGTGATGATGCAATG 260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 376 CTGTACATGGCCACAGCCACCCCTGTGTGATGATGCAATG 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 7  
US-09-918-995-6557  
Sequence 6557, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918, 995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235, 076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6557  
LENGTH: 417  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)-(417)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-6557

Query Match  
Best Local Similarity 20.7%; Score 171.8; DB 9; Length 417;  
Matches 173; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
OY 41 CCTGACAGCCCTGTGTTCCCAAGCAGAGAAATACCTGTGAGCCCTCTTCATA 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 242 CCGGCGACCCCTGTGTTCCCAAGCAGAGAAATACCTGTGAGCCCTCTTCATA 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 101 GAACAGAGATGATGCTGTGATTTCAAGACCTATGTGATCAGGCTGCAGAGCTGCT 160
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 302 GAACAGAGATGATGCTGTGATTTCAAGACCTATGTGATCAGGCTGCAGAGCTGCT 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 161 GAGGAGTTTGTCAATGTCTACTACACACCATGTGTAAGCGGCGCTTGTCTGT 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 362 GAGGAGTTTGTCAATGTCTACTACACACCATGTGTAAGCGGCGCTTGTCTGT 416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 8  
US-09-764-864-1682/c  
Sequence 1682, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764, 864  
CURRENT FILING DATE: 2001-01-17  
PRIOR APPLICATION data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: Patencin Ver. 2.0  
SEQ ID NO: 1682  
LENGTH: 8918  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-864-1682

Query Match  
Best Local Similarity 13.6%; Score 113; DB 10; Length 8918;  
Matches 152; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
OY 347 GTTCATGATGAAGCCACACACAGCCAGCAGCTGTTGTGATGTGATGATGCTG 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2199 GATGAAGAGCAAGCAACCTACCTCCCAACTACATGTTCTTGTGACCACTGATCTG 2140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies.  
FILE REFERENCE: PC005  
CURRENT APPLICATION NUMBER: US/09/764,877  
CURRENT FILING DATE: 2001.01.17  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 4031  
SOFTWARE: PatentIn Ver. 2.0

us-09-763-902b-10.rnpb

Page 6

```

; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Glycine max
US-09-443-704-11

```

Query Match	4.28;	Score 34.8;	DB 10;	Length 1141;
Best Local Similarity	57.38;	Pred. No. 1.5;		
Matches 63;	Conservative 0;	Mismatches 47;	Total 110;	

QY 667 GGTGAACCTCTTTTGTGTCACAGTCTAGAGTCCCTTCCGTAATAATACTGTG 726  
 614 GGTTCCATCTGTGTGTCATTAAGCACTATAGACACCAACGAAATACCTTCTTCTTTT  
 Db

OY 727 TGTCAAGTTCTTCCTTCAACGATGAACATTCTTACTATTTCTACTTGC 776  
| | | | | | | | | | | | | | | |  
Db 554 GTTGTCGATATCATTCGATCTTGTCTGTGTAATGGTAACTTTCTTCCG 505

RESULT 15  
US-10-097

Sequence 31, Application US/10097065  
Publication No. US20030055236A1  
GENERAL INFORMATION:  
APPLICANT: Moore, Paul A. et al.  
TITLE OF INVENTION: 110 Human Secreted Protein  
FILE REFERENCE: P2021P1  
CURRENT APPLICATION NUMBER: US/10/097,065  
PRIOR FILING DATE: 2002-03-14  
PRIOR APPLICATION NUMBER: PCT/US98/27059  
PRIOR FILING DATE: 1998-12-17  
PRIOR APPLICATION NUMBER: 60/070,923  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,007  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,057  
PRIOR FILING DATE: 1997-12-18  
PRIOR APPLICATION NUMBER: 60/068,006  
PRIOR FILING DATE: 1997-12-18

PRIOR FILING DATE: 1997-12-19  
PRIOR APPLICATION NUMBER: 60/068,367  
PRIOR FILING DATE: 1997-12-19

PRIOR APPLICATION NUMBER: 60/068,368  
 PRIOR FILING DATE: 1997-12-19  
 PRIOR APPLICATION NUMBER: 60/068,169  
 PRIOR FILING DATE: 1997-12-19  
 PRIOR APPLICATION NUMBER: 60/068,053  
 PRIOR FILING DATE: 1997-12-18  
 PRIOR APPLICATION NUMBER: 60/068,064  
 PRIOR FILING DATE: 1997-12-18

PRIOR FILING DATE: 1997-12-18  
 PRIOR APPLICATION NUMBER: 60/068,008  
 PRIOR FILING DATE: 1997-12-18

PRIOR APPLICATION NUMBER: 60/068,365  
PRIOR FILING DATE: 1997-12-19  
NUMBER OF SEQ ID NOS: 672

```
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 31
LENGTH: 1490
TYPE: DNA
ORGANISM: Homo sapiens
S-10-097-065-31
```

Query Match	4.28;	Score 34.8;	DB 9;	Length 1490;
Best Local Similarity	58.8%;	Pred. No. 1.8;		

	Matches	Conservative	Mismatches	Indels	Gaps
471 CCAGCACACAGTGTGGAGATCGCAGTgTCGCTTCCGCTTCAGAGACTGGGCGACT	60	0	42	0	0
530					

236 CCAGCTCCTGGGCACGGCGTTGGAGGCTGCGCTTGTGTTGACACAGAGCAGGA 177

531 AGTGGGGTGGCAGAGGCTCTTTGCTTCATTCAGCCCTAGC 572

Mon Jun 30 08:51:11 2003

us-09-763-902b-10.rnpb

Page 7

Db 176 TGAGGGGAGGAGGATTCTTACGATGACAGGCGCACCCGTAC 135

Search completed: June 28, 2003, 03:41:58  
Job time : 142.193 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:27:44 ; Search time 45.3644 Seconds  
(Without alignments)  
5611.049 Million cell updates/sec

Title: US-09-763-902B-10

Perfect score: 830

Sequence: 1 tgaacttgcatcactg.....ataattaataacacatg 830

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Issued Patents: NA:  
1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PCITUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	41.2	5.0	7218 1	US-08-232-463-14 Sequence 14, Appl
2	35	4.2	8478 3	US-08-676-818-11 Sequence 1, Appl
3	35	4.2	8478 4	US-09-407-549-1 Sequence 1, Appl
4	33.6	4.0	3957 1	US-07-689-008-5 Sequence 5, Appl
5	33.6	4.0	9540 1	US-07-689-008-5 Sequence 1, Appl
6	33.6	4.0	4403765 4	US-09-103-840A-2 Sequence 2, Appl
7	33.6	4.0	4411529 4	US-09-103-840A-1 Sequence 1, Appl
8	33	4.0	2661 4	US-09-134-001C-1659 Sequence 1659, Ap
9	32	3.9	11811 4	US-09-078-294-7 Sequence 7, Appl
10	31.4	3.8	16836 4	US-09-147-236-1 Sequence 1, Appl
11	31.4	3.8	16836 4	US-09-147-236-1 Sequence 10, Appl
12	31.2	3.8	3114 3	US-08-946-026-12 Sequence 12, Appl
13	30.8	3.7	2109 2	US-08-835-099A-5 Sequence 5, Appl
14	30.8	3.7	2109 2	US-09-157-349-5 Sequence 5, Appl
15	30.8	3.7	2136 2	US-08-835-099A-6 Sequence 6, Appl
16	30.8	3.7	2136 3	US-09-157-349-6 Sequence 6, Appl
17	30.8	3.7	2237 1	US-08-487-135B-1 Sequence 1, Appl
18	30.8	3.7	2237 2	US-08-915-972A-1 Sequence 1, Appl
19	30.8	3.7	2237 2	US-09-177-909-1 Sequence 1, Appl
20	30.6	3.7	1179 4	US-09-247-373B-43 Sequence 43, Appl
21	30.6	3.7	2277 1	US-08-676-967-2 Sequence 2, Appl
22	30.6	3.7	2277 1	US-08-676-967-2 Sequence 2, Appl
23	30.6	3.7	2277 2	US-09-098-487-2 Sequence 2, Appl
24	30.4	3.7	848 4	US-08-905-223-27 Sequence 27, Appl
25	30.4	3.7	848 4	US-09-247-155-27 Sequence 27, Appl
26	30.4	3.7	1364 1	US-08-265-087-3 Sequence 3, Appl
27	30.4	3.7	1364 1	US-08-621-493-3 Sequence 3, Appl

C 28	30.4	3.7	1364 2	US-08-965-688-3 Sequence 3, Appl
C 29	30.4	3.7	1364 4	US-09-260-173-3 Sequence 3, Appl
C 30	30.4	3.7	4601 1	US-08-470-058-3 Sequence 3, Appl
C 31	30.4	3.7	4601 3	US-09-037-188-3 Sequence 3, Appl
C 32	30.4	3.7	4601 4	US-09-285-310-3 Sequence 3, Appl
C 33	30.4	3.7	5095 1	US-08-092-817-3 Sequence 3, Appl
C 34	30.4	3.7	5095 4	US-08-485-128-3 Sequence 3, Appl
C 35	30.2	3.6	55827 4	US-09-813-133A-3 Sequence 454, App
C 36	30	3.6	565 4	US-09-328-111-454 Sequence 63, Appl
C 37	30	3.6	1114 4	US-09-378-088A-63 Sequence 10, Appl
C 38	30	3.6	1158 3	US-08-633-993A-10 Sequence 10, Appl
C 39	30	3.6	1158 4	US-08-844-188-10 Sequence 10, Appl
C 40	30	3.6	1158 4	US-09-378-088A-10 Sequence 67, Appl
C 41	30	3.6	1158 4	US-09-378-088A-67 Sequence 89, Appl
C 42	30	3.6	1186 4	US-09-378-088A-89 Sequence 11, Appl
C 43	30	3.6	1334 4	US-09-347-798-11 Sequence 30, Appl
C 44	30	3.6	2015 3	US-08-633-993A-30 Sequence 30, Appl
C 45	30	3.6	2015 3	US-08-844-188-30 Sequence 30, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14  
Query Match 5.0%; Score 41.2; DB 1; Length 7218;

1

Db 3362 CTGTGCGGCTGCTCAGGAGGCTTT

Db 3362 CTGTGGGCTGCTCAGGAGGCTTCAACATCATTTGAAGCCAGCAGGGAACGACAGCT 3421

OY 208 TTTCCTCCCGCTGTACATGGGACAGCCCGTGTGATGGCAATGCTTC 267  
DB 3422 TTTATTTTCTATATGACATGACAGACCGTGTGAATATGGTTATGTGGTGA 3481  
OY 268 AGACAGAAATCCCTTGAGT 286  
DB 3482 AGGAGATCACACACCGCAAT 3500

RESULT 4  
US-07-689-008-5/c  
Sequence 5, Application US/07689008  
Patent No. 5268274

GENERAL INFORMATION:

APPLICANT: Ben-Bassat, Arie

APPLICANT: Calhoon, Roger D

APPLICANT: Fear, Anna L

APPLICANT: Gelfand, David H

APPLICANT: Meade, James H

APPLICANT: Tal, Rony

APPLICANT: Wong, Hing

APPLICANT: Benziman, Moshe

TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE

NUMBER OF SEQUENCES: 15

EXPRESSION OF CELLULOSE SYNTHASE OPERON

CORRESPONDENCE ADDRESS:

ADDRESSEE: McCutchen, Doyle, Brown & Enersen

STREET: Three Embarcadero Center

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/689,008

FILING DATE: 19910422

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 337,194

FILING DATE: 12-APR-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 496,236

FILING DATE: 23-MAR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Murphy, Lisabeth Felix

REGISTRATION NUMBER: 31547

TELEPHONE: (415) 393-2000

TELEFAX: (415) 393-2286

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 3957 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-07-689-008-5

Query Match 4.0%; Score 33.6; DB 1; Length 3957;

Best Local Similarity 56.2%; Pred. No. 1.8; Mismatches 49; Indels 0; Gaps 0;

Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 468 CACCCAGCAACAGTGTGGAAGATCGCAAGTACTGCTTCGCTTCACAGACTGGGCA 527

DB 484 CACCGTGGAAACAGCTTCTGTAGCCGCCACCGCTGCGCGTCTGCGCCGAAACCGCA 425

OY 528 GCTAGTGGGGGTGGCAGAGGCTCTTTGCTTCATTAGCCCTAGCTCTAG 579

DB 424 GCGAGCGATGCGACAGTCCCTTGGAGATGCCCGCTCCGTGAGCAG 373

RESULT 5  
US-07-689-008-1/c  
Sequence 1, Application US/07689008  
Patent No. 5268274

GENERAL INFORMATION:

APPLICANT: Ben-Bassat, Arie

APPLICANT: Calhoon, Roger D

APPLICANT: Fear, Anna L

APPLICANT: Gelfand, David H

APPLICANT: Meade, James H

APPLICANT: Tal, Rony

APPLICANT: Wong, Hing

APPLICANT: Benziman, Moshe

TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE

NUMBER OF SEQUENCES: 15

EXPRESSION OF CELLULOSE SYNTHASE OPERON

CORRESPONDENCE ADDRESS:

ADDRESSEE: McCutchen, Doyle, Brown & Enersen

STREET: Three Embarcadero Center

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/689,008

FILING DATE: 19910422

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 337,194

FILING DATE: 12-APR-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 496,236

FILING DATE: 23-MAR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Murphy, Lisabeth Felix

REGISTRATION NUMBER: 31547

TELEPHONE: (415) 393-2000

TELEFAX: (415) 393-2286

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 9540 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: join(328..2589, 2594..4999, 5005..8961, 8964

LOCATION: 9431)

OTHER INFORMATION:

US-07-689-008-1

Query Match 4.0%; Score 33.6; DB 1; Length 9540;

Best Local Similarity 56.2%; Pred. No. 3.1; Mismatches 49; Indels 0; Gaps 0;

Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 468 CACCCAGCAACAGTGTGGAAGATCGCAAGTACTGCTTCGCTTCACAGACTGGGCA 527

DB 5488 CACCGTGGAAACAGCTTCTGTAGCCGCCACCGCTGCGCGTCTGCGCCGAAACCGCA 5429

OY 528 GCTAGTGGGGGTGGCAGAGGCTCTTTGCTTCATTAGCCCTAGCTCTAG 579

Db 5428 GCGAGCGAGTCTGCGAGCAGGTGCGCTTGGAGATGCGCCGCTGCTGAGCAG 5377

## RESULT 6

US-09-103-840A-2

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

US-09-103-840A-2

Query Match  
Best Local Similarity 4.0%; Score 33.6; DB 4; Length 4403765;  
Matches 81; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 400 ATGAGGAACTTTGAGGGGACAAACAGGAGCTTCAACCACTTCTGACCGC 459

Db 2768205 ATCGGTGGTGTGATGGCTGTATCCGCACCTGGACATCAGCAGCAGCGCCGCG 2768264

QY 460 CCAGGCGCTCACCCAGCAACAGCTGTGAAGATCGCAAGAGTCTCCGCTTCCAGCA 519

Db 2768265 CGAGGAGTTCTTACGAGCAGCCCGAGGTTCCGCCGATTCGCGCTGCGAGGG 2768324

QY 520 CTGGGCGCACTAGTGGGGGTGCGAGAGTCTCTTCTTC 559

Db 2768325 CCAGGCGCGGCTGACAGATCGTTGGGGGCCCACTTC 2768364

## RESULT 7

US-09-103-840A-1

Sequence 1, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 1

LENGTH: 4411529

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

OTHER INFORMATION: H37RV

Query Match  
Best Local Similarity 4.0%; Score 33.6; DB 4; Length 4411529;  
Matches 81; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 400 ATGAGTGAAGTTTGGAGGGACAAACAGGAGCTTCAACCACTTCTGACCGC 459

Db 2771395 ATCGGTGGTGTGATGGCTGTATCCGCACCTGGACATCAGCAGCAGCGCCGCG 2771454

QY 460 CCAGGCGCTCACCCAGCAACAGCTGTGAAGATCGCAAGAGTCTCCGCTTCCAGCA 519

Db 2771455 CGAGGAGTTCTTACGAGCAGCCCGAGGTTCCGCCGATTCGCGCTGCGAGGG 2771514

QY 520 CTGGGCGCACTAGTGGGGGTGCGAGAGTCTCTTCTTC 559

Db 2771515 CCAGGCGCGGCTGACAGATCGTTGGGGGCCCACTTC 2771554

## RESULT 8

US-09-134-001C-1659/c

Sequence 1659, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 1659

LENGTH: 2661

TYPE: DNA

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1659

Query Match  
Best Local Similarity 4.0%; Score 33; DB 4; Length 2661;  
Matches 57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 672 AACTCTTTTGTGCTCAAGTCTAGAGAGTCCCTTCCGATATATATCTGTTGCA 731

Db 2253 AACTATCATATATATATATATCTGTTGTTCTTTTCAATATGATATATTAAC 2194

QY 732 TACTTCTTTTCAAGTAACTTCTTCTATTTT 768

Db 2193 TATCTCTCTTTAAAGTGAATTTCTTCTACTTT 2157

## RESULT 9

US-09-078-294-7

Sequence 7, Application US/09078294

Patent No. 6265211

GENERAL INFORMATION:

APPLICANT: Choo, Kong-Hong Andy

APPLICANT: Du Salt, Desiree

APPLICANT: Cancilla, Michael R.

TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE

FILE REFERENCE: Davles Col

CURRENT APPLICATION NUMBER: US/09/078,294

CURRENT FILING DATE: 1998-05-13

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 7

LENGTH: 11811

TYPE: DNA

ORGANISM: BAC-12 contig 3

Query Match  
Best Local Similarity 3.9%; Score 32; DB 4; Length 11811;  
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 671 GAATCTTTTGTGCTCAAGTCTAGAGAGTCCCTTCCGATATATATATCTGTTGTC 730

DB 4379 GCACCTGGTACTGATCTTTAATTTTGTCTACTTAACTGAGGCAATATCTGTATTT 4438  
QY 731 ATGTTCTCTTCAAGTAGTAACCTTTCTATTCTTCTACTT 774  
DB 4439 ATGTTTACTTTTATTTCTTATGATATATAGATTGACGAT 4482

RESULT 10  
US-09-147-236-1/c  
Sequence 1, Application US/09147236A  
Patent No. 6316251  
GENERAL INFORMATION:

APPLICANT: TONOUCHI, Naoto  
APPLICANT: TSUCHIDA, Takayasu  
APPLICANT: YOSHINAGA, Fumihiko  
APPLICANT: TAHARA, Naoki  
APPLICANT: HAYASHI, Takahisa  
TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE  
FILE REFERENCE: 6537-011-0PCT  
CURRENT APPLICATION NUMBER: US/09/147,236A  
CURRENT FILING DATE: 1999-04-08  
EARLIER APPLICATION NUMBER: PCT/JP97/03633  
EARLIER FILING DATE: 1997-10-09  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 16836  
TYPE: DNA  
ORGANISM: Acetobacter xylinum

FEATURE:  
NAME/KEY: CDS  
LOCATION: (869)..(1891)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3101)..(5368)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (5373)..(7778)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (7784)..(11761)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (11764)..(12231)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (12448)..(14652)  
FEATURE:  
OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, t, or  
OTHER INFORMATION: c  
US-09-147-236-1

Query Match 3.8%; Score 31.4; DB 4; Length 16836;  
Best Local Similarity 54.9%; Pred. No. 22;  
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 468 CACCAGACACAGTGTGGAAGATGCAAGTACTCTTCGCTTCAGAGACTGGGCA 527  
DB 8267 CACCGTAGAAGAGCTTCTGTAACCCGCGCGCTGCGCCGGAACCGCCA 8208  
QY 528 GCTAGTGGGGTGGCAGAGGCTCTTTGCTTCACTGAGCCCTAGCTCTAGA 580  
DB 8207 GCGAAGGATCTGTGACAGAGTGTGCGTTTGGGAATGGCCGCTGAGACAGA 8155

RESULT 11  
US-09-147-236-10/c  
Sequence 10, Application US/09147236A  
Patent No. 6316251  
GENERAL INFORMATION:  
APPLICANT: TONOUCHI, Naoto  
APPLICANT: TSUCHIDA, Takayasu

APPLICANT: YOSHINAGA, Fumihiko  
APPLICANT: TAHARA, Naoki  
APPLICANT: HAYASHI, Takahisa  
TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE  
FILE REFERENCE: 6537-011-0PCT  
CURRENT APPLICATION NUMBER: US/09/147,236A  
CURRENT FILING DATE: 1999-04-08  
EARLIER APPLICATION NUMBER: PCT/JP97/03633  
EARLIER FILING DATE: 1997-10-09  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 10  
LENGTH: 16836  
TYPE: DNA  
ORGANISM: Acetobacter xylinum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1891)..(2922)  
FEATURE:  
OTHER INFORMATION: Nucleotide sequence is the same as SEQ ID NO:1  
OTHER INFORMATION: t  
US-09-147-236-10

Query Match 3.8%; Score 31.4; DB 4; Length 16836;  
Best Local Similarity 54.9%; Pred. No. 22;  
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 468 CACCAGACACAGTGTGGAAGATGCAAGTACTCTTCGCTTCAGAGACTGGGCA 527  
DB 8267 CACCGTAGAAGAGCTTCTGTAACCCGCGCGCTGCGCCGGAACCGCCA 8208  
QY 528 GCTAGTGGGGTGGCAGAGGCTCTTTGCTTCACTGAGCCCTAGCTCTAGA 580  
DB 8207 GCGAAGGATCTGTGACAGAGTGTGCGTTTGGGAATGGCCGCTGAGACAGA 8155

RESULT 12  
US-08-946-026-12  
Sequence 12, Application US/08946026  
Patent No. 6034218  
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.  
APPLICANT: Dillion, Davin C.  
APPLICANT: Twartzik, Daniel R.  
APPLICANT: Mitcham, Jennifer L.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/946,026  
FILING DATE: 07-OCT-1997  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
NAME: MAKI, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.424C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031



INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3114 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-946-026-12

Query Match  
Best Local Similarity 58.7%; Score 31.2; DB 3; Length 3114;  
Matches 54; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 204 GGGGTTGCTGCTCCCGCTGATCGGCACACCCCTGGTCTGGAATGCAATGCTG 263  
DB 2581 GTGTGTTCTGCTGCTCCCTGCTGAGGCTATCTGAGGGTGAACCTGCAATTCG 2640  
QY 264 TTTCAGGACAGAACTCTGAGTGAAGTTT 295  
DB 2641 TTTCCTTGTAAATATTTGCTGCTCTTT 2672

## RESULT 13

US-08-835-099A-5  
Sequence 5, Application US/08835099A  
Patent No. 5874277

## GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi  
APPLICANT: NISHI, Kazuo  
APPLICANT: KANAMOTO, Tomohiro  
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION  
TITLE OF INVENTION: AND USE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/835,099A  
FILING DATE: 04-APR-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 083649/1996  
FILING DATE: 05-APR-1996  
APPLICATION NUMBER: 97105508.2  
FILING DATE: 03-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 47342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX:

## INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2109 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-835-099A-5

Query Match  
Best Local Similarity 55.7%; Score 30.8; DB 2; Length 2109;  
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 374 ACCAGGTCCTGTTGTCATCTGTGATCATGTAAGTTGAGGAGCAACAAACAGGAC 433  
DB 1288 AGCATGGCTATGCGGCTACACAGGTTCACAGAGAGTGAGATCAGAGGACGACAC 1347  
QY 434 TTCAACGACAGACTCATCTGACCGCCAGGCTCAGCCAGCAACA 479  
DB 1348 TTGGGCGGGATTTCTTCTGCGCTACACAGCCCTCAGCCGACCA 1393

## RESULT 14

US-09-157-349-5  
Sequence 5, Application US/09157349  
Patent No. 6068990

## GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi  
APPLICANT: NISHI, Kazuo  
APPLICANT: KANAMOTO, Tomohiro  
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION  
TITLE OF INVENTION: AND USE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/157,349  
FILING DATE:

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/835,099  
FILING DATE:  
APPLICATION NUMBER: 97105508.2  
FILING DATE: 03-APR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 47342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX:

## INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2109 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-157-349-5

Query Match  
Best Local Similarity 55.7%; Score 30.8; DB 3; Length 2109;  
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 374 ACCAGGTCCTGTTGTCATCTGTGATCATGTAAGTTGAGGAGCAACAAACAGGAC 433  
DB 1288 AGCATGGCTATGCGGCTACACAGGTTCACAGAGAGTGAGATCAGAGGACGACAC 1347  
QY 434 TTCAACGACAGACTCATCTGACCGCCAGGCTCAGCCAGCAACA 479  
DB 1348 TTGGGCGGGATTTCTTCTGCGCTACACAGCCCTCAGCCGACCA 1393

## RESULT 15

US-08-835-099A-6

; Sequence 6, Application us/08835099A

; Patent No. 5874277

GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi

APPLICANT: NISHI, Kazuo

APPLICANT: KANAMOTO, Tomohiro

TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION

TITLE OF INVENTION: AND USE

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/835,099A

FILING DATE: 04-APR-1997

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 083649/1996

FILING DATE: 05-APR-1996

APPLICATION NUMBER: 97105508.2

FILING DATE: 03-APR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Resnick, David S

REGISTRATION NUMBER: 34,235

REFERENCE/DOCKET NUMBER: 47342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2136 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-835-099A-6

Query Match 3.7%; Score 30.8; DB 2; Length 2136;

Best Local Similarity 55.7%; Pred. NO. 10;

Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

OY	374	ACCAAGGTCCTGTTGTCATCTGTGATCACTGAAGTTTGAGGGGACAAACAGCGGAC	433
DB	1315	AGCATCGGCTATGCGCTGTACAGAGTCCCAAGAGCTGAGAGTCAACAGGACGAC	1374
OY	434	TTCAACCAAGAACTCATCTGACCGCCAGGCTCACCAGACACA	479
DB	1375	TTGGCCGGGATTTCTCTGCTACAGCCCTCAGCCGACCA	1420

Search completed: June 28, 2003, 03:36:22

Job time : 53.3644 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:22:51 ; Search time 224.7 Seconds

(without alignments)  
10761.067 Million cell updates/sec

Title: US-09-763-902b-10

Sequence: 1 tgaacttgcatcaccgctg.....ataattataataacacg 830

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank:  
1: gb\_ba:\*  
2: gb\_bt:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_ov:\*  
7: gb\_ov:\*  
8: gb\_ov:\*  
9: gb\_ov:\*  
10: gb\_ov:\*  
11: gb\_ov:\*  
12: gb\_ov:\*  
13: gb\_ov:\*  
14: gb\_ov:\*  
15: gb\_ov:\*  
16: gb\_ov:\*  
17: gb\_ov:\*  
18: gb\_ov:\*  
19: gb\_ov:\*  
20: gb\_ov:\*  
21: gb\_ov:\*  
22: gb\_ov:\*  
23: gb\_ov:\*  
24: gb\_ov:\*  
25: gb\_ov:\*  
26: gb\_ov:\*  
27: gb\_ov:\*  
28: gb\_ov:\*  
29: gb\_ov:\*  
30: gb\_ov:\*  
31: gb\_ov:\*  
32: gb\_ov:\*  
33: gb\_ov:\*  
34: gb\_ov:\*  
35: gb\_ov:\*  
36: gb\_ov:\*  
37: gb\_ov:\*  
38: gb\_ov:\*  
39: gb\_ov:\*  
40: gb\_ov:\*  
41: gb\_ov:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	830	100.0	112180	9	HSJ322G13	AL096677 Human DNA
2	786.8	94.8	1176	9	BC003410	BC003410 Homo sapi
3	759.8	91.5	965	9	AF156957	AF156957 Homo sapi
4	736.4	88.7	939	9	BC002687	BC002687 Homo sapi
5	733.2	88.3	963	9	BC003029	BC003029 Homo sapi
6	733.2	88.3	972	9	BC000759	BC000759 Homo sapi
7	719.2	86.7	1049	9	AK026360	AK026360 Homo sapi
8	508.4	61.3	135619	9	AL354809	AL354809 Homo sapi
9	508.4	61.3	147310	2	AL513495	AL513495 Homo sapi
10	492	59.3	186415	2	AC110699	AC110699 Rattus no
11	381.2	45.9	644	10	AF156958	AF156958 Mus muscu
12	357.8	43.1	182118	2	AC128310	AC128310 Rattus no
13	317	38.2	318	6	AX261550	AX261550 Sequence
14	287.6	34.7	153483	2	AC111510	AC111510 Rattus no
15	251.6	30.3	896	9	AF201942	AF201942 Homo sapi
16	251.6	30.3	2577	9	BC014888	BC014888 Homo sapi
17	251.6	30.3	2692	9	AK023289	AK023289 Homo sapi
18	250.2	30.1	429	9	AF246127	AF246127 Homo sapi
19	250.2	30.1	429	9	HS277591	HS277591 Homo sapi
20	248.6	30.0	567	9	HS277591	HS277591 Homo sapi
21	223	26.9	247	6	AX330381	AX330381 Sequence
22	132	15.9	133	11	G42938	G42938 WtAF-410-ST
23	113	13.6	17863	9	HS596C15	AL031387 Human DNA
24	105	12.7	148320	2	AC107438	AC107438 Rattus no
25	105	12.7	173939	2	AC110357	AC110357 Rattus no
26	105	12.7	186222	2	AC123243	AC123243 Rattus no
27	105	12.7	232171	10	AL731672	AL731672 Mouse DNA
28	81.2	9.8	238817	2	AC102111	AC102111 Mus muscu
29	76.4	9.2	507	3	AF156959	AF156959 Drosophill
30	76.4	9.2	538	3	AF071250	AF071250 Drosophill
31	59.4	7.2	295166	2	AL161914	AL161914 Homo sapi
32	58.8	7.1	42244	2	AC020482	AC020482 Drosophill
33	58.8	7.1	147086	3	AC009182	AC009182 Drosophill
34	58.8	7.1	175118	3	AC010842	AC010842 Drosophill
35	58.8	7.1	303823	3	AE003462	AE003462 Drosophill
36	57.6	6.9	173939	2	AC110357	AC110357 Rattus no
37	56	6.7	203822	2	AC073698	AC073698 Mus muscu
38	48.6	5.9	207621	2	AL845495	AL845495 Mus muscu
39	47.4	5.7	175991	2	AC094606	AC094606 Rattus no
40	41.6	5.0	38979	2	AC104197	AC104197 Mus muscu
41	41.2	5.0	7218	6	166494	166494 Sequence 14
42	41.2	5.0	110000	9	AL355364	AL355364 Homo sapi
43	41.2	5.0	125020	9	AF429315	AF429315 Homo sapi
44	41.2	5.0	139325	9	AL589943	AL589943 Human DNA
45	41	4.9	215219	2	AC118225	AC118225 Mus muscu

## ALIGNMENTS

RESULT 1  
HSJ322G13  
LOCUS  
DEFINITION  
Human DNA sequence from clone RP3-322G13 on chromosome 20p11.21-12.3 Contains the gene for NTF2-related export protein (NXT1), a gene for zinc finger protein FLJ21794, two putative novel genes, a gene for beta-soluble NSF attachment protein (SNAP-beta), a novel gene for a protein similar to cystatin, another novel gene for a protein similar to cystatin 8 (CST8), ESTs, STSs, GSSs and six Cpg islands, complete sequence.  
ACCESSION  
AL096677  
VERSION  
AL096677.21 GI:10862836  
KEYWORDS  
HTG; Cpg island; CST8; cystatin; export protein; FLJ21794; NSF attachment protein; NTF2; NXT1; SNAP; zinc finger protein.  
SOURCE  
human.



```

repeat_region 15090..15383
/note="AluX repeat: matches 1..294 of consensus"
15384..15518
/note="AluX repeat: matches 1..138 of consensus"
15519..15698
/note="L2 repeat: matches 2709..2750 of consensus"
15714..16375
/note="MER67C repeat: matches 1..710 of consensus"
16723..17983
/note="CpG Island"
/evidence-not-experimental
repeat_region 18803..18837
/note="MIR repeat: matches 126..160 of consensus"
complement(18817..19228)
/note="match: GSS: Em:AQ203877"
complement(18864..19228)
/note="match: GSS: Em:AQ145317"
18914..18973
/note="MIR repeat: matches 74..140 of consensus"
19249..19878
/note="match: GSS: Em:AQ895221"
19598..28277
/gene="d3j22613.2"
join(19598..20961,22217..22311,23976..24143,24798..24955,
25305..28277)
/gene="d3j22613.2"
/product="d3j22613.2.3 (zinc finger protein FLJ21794,
isoform 3)"
/note="match: CDNA3: Em:AL110217 Em:AF167320 Em:X78927
Em:AJ224763 Em:AL080123 Em:AF161544 Em:Z67747 Em:AB037770
Em:X60155 Em:X82125 Em:U66561 Em:AF022146 Em:L36315
Em:M36514 Em:X12592 Em:X64413 Em:AF022158 Em:L75847
Em:DI0632 Em:X06021 Em:X89264 Em:X78933 Em:L5309
Em:X52533 Em:S67970 Em:AK025447
match: ESTs: Em:AA17107 Em:R68950 Em:AT023363 Em:AI217180
Em:AM291034 Em:AA17012 Em:AA701588 Em:AI177089
Em:AI145986 Em:AW079910 Em:AI961159 Em:AW019926
Em:AA035213 Em:AW082641 Em:AW594742 Em:AW531679
Em:AI984655 Em:AA273573 Em:AA173862 Em:AV236342
Em:AW144279 Em:AL135710 Em:AA365129 Em:AA346014
Em:AA062926 Em:AL046380 Em:AA254007 Em:T85404 Em:AA383356

Query Match 100.0% Score 830 DB 9 Length 112180
Best Local Similarity 100.0% Pred. No. 4.7e-225 Indels 0 Gaps 0
Matches 830 Conservative 0 Mismatches 0

QY 1 TGAACCTTGGCATTCACGTCGCTTCTTCAACCTTACTCCCTGCAGCCCGGTGCC 60
|||||
Db 9147 TGAACCTTGGCATTCACGTCGCTTCTTCAACCTTACTCCCTGCAGCCCGGTGCC 9206

QY 61 CAAGCAGAGAAATACCTGTGTGAGACCCCTCTTCATGAACAGAGATGCATCTGT 120
|||||
Db 9207 CAAGCAGAGAAATACCTGTGTGAGACCCCTCTTCATGAACAGAGATGCATCTGT 9266

QY 121 GGAATTCAGAGACTAGTGTGATCAGACCTGCAGAGCTGCTGAGAGATTTCATGCTCTA 180
|||||
Db 9267 GGAATTCAGAGACTAGTGTGATCAGACCTGCAGAGCTGCTGAGAGATTTCATGCTCTA 9326

QY 181 CTACACCAACCATGATTAAGCGGGCGGCTTGTCTCCCGCTGTACATGGGACAGCCAC 240
|||||
Db 9327 CTACACCAACCATGATTAAGCGGGCGGCTTGTCTCCCGCTGTACATGGGACAGCCAC 9386

QY 241 CCTGGTCTGGAATGGCAATGCTTTTCAGAGCAAGAATCCTTGAAGTATTTTGAAT 300
|||||
Db 9387 CCTGGTCTGGAATGGCAATGCTTTTCAGAGCAAGAATCCTTGAAGTATTTTGAAT 9446

QY 301 GTTGCCTTCCAGGAGTTCCTCAATACGCGTGTAGACCTGCACCTGTTCATGATGAAC 360
|||||
Db 9447 GTTGCCTTCCAGGAGTTCCTCAATACGCGTGTAGACCTGCACCTGTTCATGATGAAC 9506

QY 361 CACACCAACCAACAGAGGCTCTTGTGTGATCTGTGATGATGAGTTTGAAGGGAA 420
|||||
Db 9507 CACACCAACCAACAGAGGCTCTTGTGTGATCTGTGATGATGAGTTTGAAGGGAA 9566

```

```

QY 421 CAACCAAGGGAGCTTCAACCAAGAACTTCATCTGACCGCCAGGCTCACCACCAACAC 480
|||||
Db 9567 CAACCAAGGGAGCTTCAACCAAGAACTTCATCTGACCGCCAGGCTCACCACCAACAC 9626

QY 481 AGTGTGAAGATGCCAGTACCTCTCCGCTTCAGAGACTGGCCAGCTAGTGGGGTG 540
|||||
Db 9627 AGTGTGAAGATGCCAGTACCTCTCCGCTTCAGAGACTGGCCAGCTAGTGGGGTG 9686

QY 541 GCAGAGCTCTCTTGTGCTTCATTCAGCCCTAGCTCTGTAGAGAAATGCAACCTGCATCT 600
|||||
Db 9687 GCAGAGCTCTCTTGTGCTTCATTCAGCCCTAGCTCTGTAGAGAAATGCAACCTGCATCT 9746

QY 601 CAAGATGTAGGAACAACAAGTTCATTTCTGTTGTCGAGACATGCAGACTCCACTG 660
|||||
Db 9747 CAAGATGTAGGAACAACAAGTTCATTTCTGTTGTCGAGACATGCAGACTCCACTG 9806

QY 661 TGGCGAGGTGAACTCTTTTGTGTGCTCAAGTCTAGAGAGTCCCTTCTGATATATA 720
|||||
Db 9807 TGGCGAGGTGAACTCTTTTGTGTGCTCAAGTCTAGAGAGTCCCTTCTGATATATA 9866

QY 721 CTGTGTTGATAGTTTCTTTTCAAGTAGTAACCTTTCTATTTTCTACTGCCCAG 780
|||||
Db 9867 CTGTGTTGATAGTTTCTTTTCAAGTAGTAACCTTTCTATTTTCTACTGCCCAG 9926

QY 781 TAGAGACTGTGATTCGAAATTCGTGACAAATATTAATTAATACATG 830
|||||
Db 9927 TAGAGACTGTGATTCGAAATTCGTGACAAATATTAATTAATACATG 9976

RESULT 2
BC003410 1176 bp. mRNA linear PRI 12-JUN-2001
DEFINITION Homo sapiens, NFE2-related export protein 1, clone MGC:4978
LOCUS BC003410.1 GI:13097317
ACCESSION BC003410
VERSION BC003410.1
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE NIH-MGC Project URL: http://mgc.ncl.nih.gov
AUTHORS Mammalla, Euthenia; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 1176)
JOURNAL Direct Submission
Submitted (20-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT NIH-MGC Project URL: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@cmm.tmc.edu.
Villalona, D.K., Luna, R.A., Hale, S.M., Huliy, S., Lu, X., Garcia,
A.M., Hollway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IRAC Plate: 3 Row: P Column: 18
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gl: 7019470.
FEATURES
location/Qualifiers
1..1176
/organism="Homo sapiens"
/db_xref="Locusid:29107"

```

/db\_xref="taxon:9606"  
 /clone="MGC:4978 IMAGE:3450767"  
 /tissue\_type="Placenta, chorioncarcinoma"  
 /clone\_id="NH\_MGC\_10"  
 /lab\_host="DH10B"  
 /note="vector: pcMV-SPOB6"  
 388. 810  
 /codon\_start=1  
 /product="NTP2-related export protein 1"  
 /protein\_id="AAH03410.1"  
 /db\_xref="GI:13097318"  
 /translation="MAYDEKTYVDQACRAAEFVNYYTMDKRRRLSLRYMGAT  
 LVMGNNAVSGEISLEFEMLPSEFQISVDCQPVHDEATPSQTTVLVYIGSVKFE  
 GNKORDFNFIITAOASPSNTVMKIASDCRFQDMS"  
 BASE COUNT 294 a 309 c 316 g 257 t  
 ORIGIN  
 Query Match 94.8%; Score 786.8; DB 9; Length 1176;  
 Best Local Similarity 99.7%; Pred. No. 6.5e-213;  
 Matches 788; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 41 CCTTCACAGCCCTGTTCCCAAGCAGAGAAATACCTGGTGGAGCCCTCTCCATA 100  
 DB 319 CCGCCACAGCCCTGTTCCCAAGCAGAGAAATACCTGGTGGAGCCCTCTCCATA 378  
 QY 101 GAACACAGATGGCATCTGTGATTTCAAGACCTATGTGATCAGGCTCGACAGCTGCT 160  
 DB 379 GAACACAGATGGCATCTGTGATTTCAAGACCTATGTGATCAGGCTCGACAGCTGCT 438  
 QY 161 GAGGAGTTTGTCAATGTCTACTACACCAACATGATGAGCGCGCGTGTGCTCCGC 220  
 DB 439 GAGGAGTTTGTCAATGTCTACTACACCAACATGATGAGCGCGCGTGTGCTCCGC 498  
 QY 221 CTGTACATGGGACACACCCCTGCTGGAATGCAATGCTGTTTCAGAGCAAGATCC 280  
 DB 499 CTGTACATGGGACACACCCCTGCTGGAATGCAATGCTGTTTCAGAGCAAGATCC 558  
 QY 281 TTGAGTGAATTTTGAATGTGCTCCACAGAGTTCGAATCAGGCTGTAGACATCC 340  
 DB 559 TTGAGTGAATTTTGAATGTGCTCCACAGAGTTCGAATCAGGCTGTAGACATCC 618  
 QY 499 CTGTACATGGGACACACCCCTGCTGGAATGCAATGCTGTTTCAGAGCAAGATCC 558  
 DB 281 TTGAGTGAATTTTGAATGTGCTCCACAGAGTTCGAATCAGGCTGTAGACATCC 340  
 QY 559 TTGAGTGAATTTTGAATGTGCTCCACAGAGTTCGAATCAGGCTGTAGACATCC 618  
 DB 499 CTGTACATGGGACACACCCCTGCTGGAATGCAATGCTGTTTCAGAGCAAGATCC 558  
 QY 341 CAGCCTGTTCATGATGAAAGCCACACACAGCAGCAGCTGCTGTTGTCATCTGTGA 400  
 DB 619 CAGCCTGTTCATGATGAAAGCCACACACAGCAGCAGCTGCTGTTGTCATCTGTGA 678  
 QY 401 TCAGTGAATTTGAGGGGAAACAACAGCGGACTTCACAGAACTTCTCTGACCGCC 460  
 DB 679 TCAGTGAATTTGAGGGGAAACAACAGCGGACTTCACAGAACTTCTCTGACCGCC 738  
 QY 461 CAGGCTGTCCAGACACAGCAGTGGAAAGTGGAAAGTGGCTCCGCTCCAGAGAC 520  
 DB 739 CAGGCTGTCCAGACACAGCAGTGGAAAGTGGAAAGTGGCTCCGCTCCAGAGAC 798  
 QY 521 TGGGCGAGTGTGGGGGTGGGAGAGTCTCTTGGTTCATTCAGCCCTAGCTGTAGA 580  
 DB 799 TGGGCGAGTGTGGGGGTGGGAGAGTCTCTTGGTTCATTCAGCCCTAGCTGTAGA 858  
 QY 581 GAAATGCAAACTCGACTCTCAAGAGTGTGAGAAACAAAGTTCATTTCTGTTGGGG 640  
 DB 859 GAAATGCAAACTCGACTCTCAAGAGTGTGAGAAACAAAGTTCATTTCTGTTGGGG 918  
 QY 641 AAGACGTGAGACTCCAGTGGCCAGAGTGAACCTTTTGGTGGCTCAATTCAGAGA 700  
 DB 919 AAGACGTGAGACTCCAGTGGCCAGAGTGAACCTTTTGGTGGCTCAATTCAGAGA 978  
 QY 701 GTCCCTTCTCAATATATATCTGTTGTGATAGTTTCTTTTCAAGAGTAACTTTT 760  
 DB 979 GTCCCTTCTCAATATATATCTGTTGTGATAGTTTCTTTTCAAGAGTAACTTTT 1038  
 QY 761 CTAATTTTCTACTGCGCCAGTAGAGACTGTGATTCGGAATTCGACAAATTAATTA 820  
 DB 1039 CTAATTTTCTACTGCGCCAGTAGAGACTGTGATTCGGAATTCGACAAATTAATTA 1098

QY 821 AATACACATG 830  
 DB 1099 AATACACATG 1108  
 RESULT 3  
 AF156957  
 LOCUS  
 DEFINITION Homo sapiens NTP2-related export protein Nxt1 (Nxt1) mRNA, complete cds.  
 ACCESSION AF156957  
 VERSION AF156957.1  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 965)  
 Black, B.E., Levesque, L., Holsa, J.M., Wood, T.C. and Paschal, B.M.  
 Identification of an NTP2-related factor that binds Ran-GTP and  
 regulates nuclear protein export  
 Mol. Cell. Biol. 19 (12), 8616-8624 (1999)  
 20036817  
 PUBMED 10567585  
 REFERENCE 2 (bases 1 to 965)  
 Black, B.E. and Paschal, B.M.  
 Direct Submission  
 Submitted (09-JUN-1999) Center for Cell Signaling, University of  
 Virginia, 7161 Hospital West, Box 577 Health Sciences Center,  
 Charlottesville, VA 22908, USA  
 FEATURES  
 source  
 1. 965  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1. 965  
 /gene="NXT1"  
 272. 694  
 /function="stimulates nuclear protein export"  
 /note="similar to the sequence presented in Genbank  
 accession Number AA85905; nuclear transport factor; binds  
 Ran-GTP"  
 /codon\_start=1  
 /product="NTP2-related export protein Nxt1"  
 /protein\_id="AA054942.1"  
 /db\_xref="GI:5880865"  
 /translation="MAYDEKTYVDQACRAAEFVNYYTMDKRRRLSLRYMGAT  
 LVMGNNAVSGEISLEFEMLPSEFQISVDCQPVHDEATPSQTTVLVYIGSVKFE  
 GNKORDFNFIITAOASPSNTVMKIASDCRFQDMS"  
 BASE COUNT 207 a 265 c 263 g 230 t  
 ORIGIN  
 Query Match 91.5%; Score 759.8; DB 9; Length 965;  
 Best Local Similarity 99.7%; Pred. No. 3.1e-205;  
 Matches 761; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 41 CCTTCACAGCCCTGTTCCCAAGCAGAGAAATACCTGGTGGAGCCCTCTCCATA 100  
 DB 203 CCGCCACAGCCCTGTTCCCAAGCAGAGAAATACCTGGTGGAGCCCTCTCCATA 262  
 QY 101 GAACACAGATGGCATCTGTGATTTCAAGACCTATGTGATCAGGCTCGACAGCTGCT 160  
 DB 263 GAACACAGATGGCATCTGTGATTTCAAGACCTATGTGATCAGGCTCGACAGCTGCT 322  
 QY 161 GAGGAGTTTGTCAATGTCTACTACACCAACATGATGAGCGCGCGTGTGCTCCGC 220  
 DB 323 GAGGAGTTTGTCAATGTCTACTACACCAACATGATGAGCGCGCGTGTGCTCCGC 382  
 QY 221 CTGTACATGGGACACACCCCTGCTGGAATGCAATGCTGTTTCAGAGCAAGATCC 280  
 DB 383 CTGTACATGGGACACACCCCTGCTGGAATGCAATGCTGTTTCAGAGCAAGATCC 442  
 QY 281 TTGAGTGAATTTTGAATGTGCTCCACAGAGTTCGAATCAGGCTGTAGACATCC 340

```

Db      443 TTGATGAGTGTGTTTGAATGTTGCTTCCAGCGAGTTCCAAATCAGCGTGTAGACTGC 502
Oy      341 CAGCGCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 400
Db      503 CAGCGCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 562
Oy      401 TCAGTGAAGTTGAGGGGAGCAACAGCGGACTTCAACAGACTTATCTGTGACGCC 460
Db      563 TCAGTGAAGTTGAGGGGAGCAACAGCGGACTTCAACAGACTTATCTGTGACGCC 622
Oy      461 CAGCGCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 520
Db      623 CAGCGCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 682
Oy      521 TGGGCGACCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCAGCCCTAGCTGTAGA 580
Db      683 TGGGCGACCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCAGCCCTAGCTGTAGA 742
Oy      581 GAAATGCAAACTGCTCAGAGATGTGAGAGACAGTTCATCTGTGTGGG 640
Db      743 GAAATGCAAACTGCTCAGAGATGTGAGAGACAGTTCATCTGTGTGGG 802
Oy      641 AGACACTGCAAGCTCAGCTGCTGCGAGGTGAACTCTTTTGTCTCAAGTTCTAGA 700
Db      803 AGACACTGCAAGCTCAGCTGCTGCGAGGTGAACTCTTTTGTCTCAAGTTCTAGA 862
Oy      701 GTCCCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 760
Db      863 GTCCCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 922
Oy      761 CTATTTTCTACTGCTCCAGTAGAGACCTGATTTCTGGAATT 803
Db      923 CTATTTTCTACTGCTCCAGTAGAGACCTGATTTCTGGAATT 965

```

```

RESULT 4
BC002687      939 bp      mRNA      linear      PRI 12-JUL-2001
DEFINITION    Homo sapiens, NMF2-related export protein 1, clone MGC:3469
IMAGE:3608037, mRNA, complete cds.
ACCESSION    BC002687
VERSION      BC002687.1 GI:12803702
KEYWORDS
SOURCE       MGC.
ORGANISM     Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 939)
Strausberg, R.
Direct Submission
Submitted (05-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA.
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue: Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
http://mgc.ncl.nih.gov/
Contact: nisc.mgc@nih.gov
Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlin, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McGowan, J., Pearson, R., Snyder, B., Stantirpop, S., Thomas, P.J.,
Tjongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

```

```

FEATURES
source
1..939
/organism="Homo sapiens"
/db_xref="LocusID:29107"
/db_xref="taxon:9606"
/clone="MGC:3469 IMAGE:3608037"
/tissue_type="uterus, endometrium adenocarcinoma"
/clone_id="NIH_MGC_44"
/lab_host="DH10B-R"
/notes="vector: pORF7"
252..674
/codon_start=1
/product="NMF2-related export protein 1"
/protein_id="AA02687.1"
/db_xref="GI:12803703"
/translation="MAYDFKTYVDQARAAEFVNVYTTTDRKRLRLSLYMGAT
LVWNGNAVSGESLSEFEMLPSEFQISVDCCPVHDEATPSQTTVLVLCGSKFE
GNKORDFNFIITROASPSMTVMKIASDPRFQDMAS"
BASE COUNT      217 a      253 c      250 g      219 t
ORIGIN
Query Match      88.7%; Score 736.4; DB 9; Length 939;
Best Local Similarity 99.2%; Pred. No. 1.5e-198;
Matches 740; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Oy      41 CCTCAGAGCCCTGCTGCTCCCAAGCAGAGAAATACCTGTGGAGCCCTCTTCATA 100
Db      183 CCGCCAGAGCCCTGCTGCTCCCAAGCAGAGAAATACCTGTGGAGCCCTCTTCATA 242
Oy      101 GAACAGAGAGTGCATCTGTGATTTCAAGACCTATGTGATCAGAGCCCTGAGACTGCT 160
Db      243 GAACAGAGAGTGCATCTGTGATTTCAAGACCTATGTGATCAGAGCCCTGAGACTGCT 302
Oy      161 GAGGAGTTTGTCAAGTCTACTACACACACATGATTAAGCGGCGGCTTGTGTCGCCG 220
Db      303 GAGGAGTTTGTCAAGTCTACTACACACACATGATTAAGCGGCGGCTTGTGTCGCCG 362
Oy      221 CTGTACATGAGGACAGCCACCTGCTGTGAATGGCAATGCTGTTCAAGACAGATCC 280
Db      363 CTGTACATGAGGACAGCCACCTGCTGTGAATGGCAATGCTGTTCAAGACAGATCC 422
Oy      281 TTGATGAGTTTGTGAATGTGCTTCCACGAGTTCCAAATACGCTGTAGACTGC 340
Db      423 TTGATGAGTTTGTGAATGTGCTTCCACGAGTTCCAAATACGCTGTAGACTGC 482
Oy      341 CAGCGCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 400
Db      483 CAGCGCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 542
Oy      401 TCAGTGAAGTTGAGGGGAGCAACAGCGGACTTCAACAGACTTATCTGTGACGCC 460
Db      543 TCAGTGAAGTTGAGGGGAGCAACAGCGGACTTCAACAGACTTATCTGTGACGCC 602
Oy      461 CAGCGCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 520
Db      603 CAGCGCTTTCATGATGAAAGCCACACAGCCAGCAGCGTCTGTGTCTATCTGTGA 662
Oy      521 TGGGCGACCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCAGCCCTAGCTGTAGA 580
Db      663 TGGGCGACCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCAGCCCTAGCTGTAGA 722
Oy      581 GAAATGCAAACTGCTCAGAGATGTGAGAGACAGTTCATCTGTGTGGG 640
Db      723 GAAATGCAAACTGCTCAGAGATGTGAGAGACAGTTCATCTGTGTGGG 782
Oy      641 AGACACTGCAAGCTCAGCTGCTGCGAGGTGAACTCTTTTGTGCTCAAGTTCTAGA 700

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 12 Row: m Column: 11  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7019470.

Db 783 AGACACTGACAGACTCCAGCTGCGGAGGTGAACTCTTTTGTGCTCAAGTCTAGGA 842  
Qy 701 GTCCCTTCTGTAATATATATCTGTTGTCATAGTTCTTCCATCAAGTAACTTTT 760  
Db 843 GTCCCTTCTGTAATATATATCTGTTGTCATAGTTCTTCCATCAAGTAACTTTT 902  
Qy 761 CTAATTTTCTACTGCGCCAGTAGAGA 786  
Db 903 CTAATTTTCTACTGCGCCAGTAGAGA 928

RESULT 5  
BC003029  
LOCUS  
DEFINITION Homo sapiens, NTF2-related export protein 1, clone MGC:4329  
IMAGE:2820775, mRNA, complete cds.  
ACCESSION BC003029.1 GI:12804338  
VERSION BC003029.1  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (06-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
REMARK Contact: MGC help desk  
COMMENT Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: DCTD/DRP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@gscc.bc.ca](mailto:info@gscc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Rehdan Chiu, Chris Fjell, Erin Garland, Ran Guln,  
Leticia Hsiao, Martin Kirywiniski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McIlvay, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prahnu, Parvaneh Saeedi, Jacqueline  
Scheil, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Scott,  
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAL Plate: 9 Row: P Column: 1  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 7019470.  
Location/Qualifiers  
1. 963  
/organism="Homo sapiens"  
/db\_xref="locusid:29107"  
/db\_xref="taxon:9606"  
/clone="MGC:4329 IMAGE:2820775"  
/tissue\_type="Lung, small cell carcinoma"  
/clone\_id="N1H\_MGC\_7"  
/lab\_host="DH10B-R"  
/note="Vector: pOT7"  
276. 698  
/codon\_start=1  
/product="NTF2-related export protein 1"  
/protein\_id="AAH03029.1"  
/db\_xref="gi:12804339"  
/translation="MAVDKTYVDACRAAEFVNYVTMRRRLSLRYMGAT  
LVNNGNVAQGESISEFEKMPSEFISIVYDQPVDEATPSTTVLVVYICGVKE  
GNQORDNQNFILTAQSPSNWTWKIASDFQDMAS"

BASE COUNT 220 a 263 c 256 g 224 t  
ORIGIN

Query Match 88.3%; Score 733.2; DB 9; Length 963;  
Best Local Similarity 98.9%; Pred. No. 1.2e-197;  
Matches 738; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 41 CCCTGCAGCCCCCTGTTGCCCAAGCAGAGAAATACCTGTGTGAGACCCCTCCATTA 100  
Db 207 CCGCGCAGCCCCCTGTTGCCCAAGCAGAGAAATACCTGTGTGAGACCCCTCCATTA 266  
Qy 101 GAACAGAGATGGCATCTGTGATTTCAAGCTATGTGATGATGAGCTCAGAGTCT 160  
Db 267 GAACAGAGATGGCATCTGTGATTTCAAGCTATGTGATGATGAGCTCAGAGTCT 326  
Qy 161 GAGAGATTTGTCAATGCTCTACTACACACCAGTGAATGAAGGGGGGCTTTGTCCCGC 220  
Db 327 GAGAGATTTGTCAATGCTCTACTACACACCAGTGAATGAAGGGGGGCTTTGTCCCGC 386  
Qy 221 CTGTACATGGGCGACAGCCACCTGTGTGATGATGAGTCTGTTCAGACAGAAATCC 280  
Db 387 CTGTACATGGGCGACAGCCACCTGTGTGATGATGAGTCTGTTCAGACAGAAATCC 446  
Qy 281 TTGAGTGAATTTTGAATGTTGCCCTTCAGGAGTTCGAAATACAGCTGTGAGATGC 340  
Db 447 TTGAGTGAATTTTGAATGTTGCCCTTCAGGAGTTCGAAATACAGCTGTGAGATGC 506  
Qy 341 CAGCCTGTTGATGATGAGACACACAGCCAGACAGCCAGGCTCTGTGTGATCTGTGA 400  
Db 507 CAGCCTGTTGATGATGAGACACACAGCCAGACAGCCAGGCTCTGTGTGATCTGTGA 566  
Qy 401 TCAGTGAATTTGAGGGGGAACAACAAGGAGCTTCAACAGAACTTCACTGACCGCC 460  
Db 567 TCAGTGAATTTGAGGGGGAACAACAAGGAGCTTCAACAGAACTTCACTGACCGCC 626  
Qy 461 CAGGCTCACCAGCAACACAGTGTGAAATGCAAGTACAGTCCGCTCCAGAGAC 520  
Db 627 CAGGCTCACCAGCAACACAGTGTGAAATGCAAGTACAGTCCGCTCCAGAGAC 686  
Qy 521 TGGGCGAGTGAATGAGGGGTGGCAGAGGTCTCTTGTGTTCAATTCAGCCCTGATGA 580  
Db 687 TGGGCGAGTGAATGAGGGGTGGCAGAGGTCTCTTGTGTTCAATTCAGCCCTGATGA 746  
Qy 581 GAAATGCAACCTGACTCTCAAGATGAGAGACCAAGTATATTTCTGTTTCCG 640  
Db 747 GAAATGCAACCTGACTCTCAAGATGAGAGACCAAGTATATTTCTGTTTCCG 806  
Qy 641 AGACACTGACAGCTCCAGCTGCGGAGGTGAATCTTTTGTGCTCAAGTCTAGGA 700  
Db 807 AGACACTGACAGCTCCAGCTGCGGAGGTGAATCTTTTGTGCTCAAGTCTAGGA 866  
Qy 701 GTCCCTTCTGTAATATATATCTGTTGTCATAGTTCTTCCATCAAGTAACTTTT 760  
Db 867 GTCCCTTCTGTAATATATATCTGTTGTCATAGTTCTTCCATCAAGTAACTTTT 926  
Qy 761 CTAATTTTCTACTGCGCCAGTAGAGA 786  
Db 927 CTAATTTTCTACTGCGCCAGTAGAGA 952

RESULT 6  
BC000759  
LOCUS  
DEFINITION Homo sapiens, NTF2-related export protein 1, clone MGC:2701  
IMAGE:2820775, mRNA, complete cds.  
ACCESSION BC000759.1 GI:12653930  
VERSION BC000759.1  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Strausberg, R.  
TITLE Direct Submission



## JOURNAL

Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA.

## REMARK

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
DNA Sequencing by: Institute for Systems Biology  
<http://www.systemsbio.org>  
contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)  
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan

## COMMENT

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>  
Series: IRAL Plate: 2 Row: 1 Column: 17  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7019470.

## FEATURES

## Source

1. 972  
Location/Qualifiers

## CDS

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:2701 IMAGE:2820775"  
/tissue="Lung, small cell carcinoma"  
/clone\_id="NIH\_MGC-7"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
285..707  
/codon\_start=1  
/product="NRF2-related export protein 1"  
/protein\_id="AAH0759.1"  
/db\_xref="GI:12653931"  
/translation="MASVDEKTYVDCAARAEEVNYVTMDKRRLLSRVWGAT  
LYWGNVYSGEISELSEFEMLPSSEKSIYVDQCPVDEATPSYTVLVYICGSVKE  
GNKRDENQFIQAQSPENYWKILASDFRQDMS"  
BASE COUNT 222 a 265 c 261 g 224 t

## ORIGIN

Query Match 88.3% Score 733.2; DB 9; Length 972;  
Best Local Similarity 98.9%; Pred. No. 1.2e-197;  
Matches 733; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

41 CCTGTGAGCCCTGGTTCCTCCAGAGAGAGAAATACCTGTGGAGCCCTCTTCATA 100  
11 |||||  
216 CCGGCGAGCCCTGGTTCCTCCAGAGAGAGAAATACCTGTGGAGCCCTCTTCATA 275  
101 GAACAGAGATGGCATGTGTGATTTCAAGACCTATGTGATCAGGCTGCAGAGCTCT 160  
|||||  
276 GAACAGAGATGGCATGTGTGATTTCAAGACCTATGTGATCAGGCTGCAGAGCTCT 335  
161 GAGAGATTTTCATATGCTACTACACCAATGATTAAGGCGCGCTTGTCTCCCGC 220  
|||||  
336 GAGAGATTTTCATATGCTACTACACCAATGATTAAGGCGCGCTTGTCTCCCGC 395  
221 CTGACATGGGACAGACCCCTGGTGTGAATGGCAATGCTTTGAGCAAGAAATCC 280  
|||||  
396 CTGACATGGGACAGACCCCTGGTGTGAATGGCAATGCTTTGAGCAAGAAATCC 455  
281 TTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 340  
|||||  
456 TTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 515  
341 CAGCCTTTCATGATGAGGACACCAAGGACGAGGAGGCTGTGTGATCTGATGGA 400  
|||||  
516 CAGCCTTTCATGATGAGGACACCAAGGACGAGGAGGCTGTGTGATCTGATGGA 575  
401 TCAGTGAAGTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 460  
|||||  
576 TCAGTGAAGTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 635

QY 461 CAGCCTTTCATGATGAGGACACCAAGGACGAGGAGGCTGTGTGATCTGATGGA 520  
|||||

DB 636 CAGCCTTTCATGATGAGGACACCAAGGACGAGGAGGCTGTGTGATCTGATGGA 695  
|||||

QY 521 TGGCGAGCTAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 580  
|||||

DB 696 TGGCGAGCTAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 755  
|||||

QY 581 GAAATGCAAACTTCGATCTTCAGAGATGTGAGGACCAAGTTCATTTCTGTGTCGG 640  
|||||

DB 756 GAAATGCAAACTTCGATCTTCAGAGATGTGAGGACCAAGTTCATTTCTGTGTCGG 815  
|||||

QY 641 AGACATCGACAGCTCCACTGTGCGAGGTGAATCTTTTGTGCTCAAGTCTAGGA 700  
|||||

DB 816 AGACATCGACAGCTCCACTGTGCGAGGTGAATCTTTTGTGCTCAAGTCTAGGA 875  
|||||

QY 701 GTCCCTTCTGATATATATCTGTTGTGATATCTTCTTTCAAAGTAGTAACCTTT 760  
|||||

DB 876 GTCCCTTCTGATATATATCTGTTGTGATATCTTCTTTCAAAGTAGTAACCTTT 935  
|||||

QY 761 CTATTTTCTACTTGGCCAGTAGGA 786  
|||||

DB 936 CTATTTTCTACTTGGCCAGTAGGA 961  
|||||

## RESULT 7

## AK026360

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## FEATURES

## Source

1. 1049  
Location/Qualifiers

## ORGANISM

/organism="Homo sapiens"  
/db\_xref="taxon:9606"

## Tissue

/tissue="Human small intestine"  
/clone\_id="HSI13259"

## Note

/note="Cloning vector pME18SFL3"  
1. 1049  
/note="Highly similar to AF156957 Homo sapiens  
NRF2-related export protein NKT1 (NKT1) mRNA"

## BASE COUNT

230 a 290 c 301 g 228 t

## ORIGIN

Query Match: 86.7%; Score 719.2; DB 9; Length 1049;  
 Best Local Similarity 98.7%; Pred. No. 1.2e-193;  
 Matches 726; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

```

QY 41 CCTGAGACCCCTGTTCCCAAGGAGGAAATACCTGTGGAGCCCTCTTCATA 100
    |||
DB 299 CCGGAGAGCCCTGTTCCCAAGGAGGAAATACCTGTGGAGCCCTCTTCATA 358
    |||
QY 101 GAACGAGATGACCTGTGGATTCAGACCTATGATGACGAGCCGACAGATGCT 160
    |||
DB 359 GAACGAGATGACCTGTGGATTCAGACCTATGATGACGAGCCGACAGATGCT 418
    |||
QY 161 GAGAGATTGTAAATGCTACTACACACCATGATGACGAGGAGGCTGTGCTCC 220
    |||
DB 419 GAGAGATTGTAAATGCTACTACACACCATGATGACGAGGAGGCTGTGCTCC 478
    |||
QY 221 CTGTACATGGGACACACCCCTGTCTGGAATGGCAATGCTGTTACAGACAAATCC 280
    |||
DB 479 CTGTACATGGGACACACCCCTGTCTGGAATGGCAATGCTGTTACAGACAAATCC 538
    |||
QY 281 TTGAGTGAATTTTGAATTTGCTTCCAGAGATTCCTCAATACGCTGTGACATCC 340
    |||
DB 539 TTGAGTGAATTTTGAATTTGCTTCCAGAGATTCCTCAATACGCTGTGACATCC 598
    |||
QY 341 CAGCCTGTATGATGATGACACACACACGAGCAGGCTGCTGTTGCTCATCTGTGA 400
    |||
DB 599 CAGCCTGTATGATGATGACACACACACGAGCAGGCTGCTGTTGCTCATCTGTGA 658
    |||
QY 401 TCAGTGAATTTGAGGGGAAACAAACAGGGACTTACACAGATTCATCTGACCGCC 460
    |||
DB 659 TCAGTGAATTTGAGGGGAAACAAACAGGGACTTACACAGATTCATCTGACCGCC 718
    |||
QY 461 CAGGCTCACCAGACACAGATGAGATGCAAGTACGCTCCGCTCCAGGAC 520
    |||
DB 719 CAGGCTCACCAGACACAGATGAGATGCAAGTACGCTCCGCTCCAGGAC 778
    |||
QY 521 TGGGCGAGCTAGTGGGGGTGGCAGAGTCTCTTCTTCACTTACGCTTCTGTAGA 580
    |||
DB 779 TGGGCGAGCTAGTGGGGGTGGCAGAGTCTCTTCTTCACTTACGCTTCTGTAGA 836
    |||
QY 581 GAATGCAAACTCTGACTCTCAAGATGTGAGGAAACAGATTCATCTGTTGTCGG 640
    |||
DB 837 GAATGCAAACTCTGACTCTCAAGATGTGAGGAAACAGATTCATCTGTTGTCGG 896
    |||
QY 641 AGACACTGAGACACTGCTGCGGAGGTGAACCTTTTGTGCTCAAGTCTAGGA 700
    |||
DB 897 AGACACTGAGACACTGCTGCGGAGGTGAACCTTTTGTGCTCAAGTCTAGGA 956
    |||
QY 701 GTCCCTTCTGTAATATATCTGTTGTCATGTTCTTTTCAAGTAGTAACCTTTT 760
    |||
DB 957 GTCCCTTCTGTAATATATCTGTTGTCATGTTCTTTTCAAGTAGTAACCTTTT 1016
    |||
QY 761 CTATTTTCTACTTGGCCAGTAGAGA 786
    |||
DB 1017 CTATTTTCTACTTGGCAAAAAAAA 1042
    |||

```

## TITLE

Direct Submission  
 Submitted (24-Feb-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Feb 28, 2001 this sequence version replaced g1:12831820:  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used to associate primary accession numbers given  
 in the feature table with their source databases: Em., EMBL; Sw.,  
 SWISSPROT; Tr., TrEMBL; Wp., WormPEP; information on the WormPEP  
 database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 13, constructed by the Sanger Centre Chromosome 13  
 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr13>  
 RP11-527F15 is from the library RP11-11.2 constructed by the group  
 of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>

## FEATURES

## source

Location/Qualifiers  
 1..135619  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="13"  
 /clone="RP11-527F15"  
 /clone\_11b="RP11-11.2"  
 55..179  
 /note="MER5B repeat: matches 42..174 of consensus"  
 396..567  
 /note="FAM repeat: matches 1..168 of consensus"  
 1056..1399  
 /note="MER7A repeat: matches 1..346 of consensus"  
 1805..2527  
 /note="L1PA16 repeat: matches 5418..6157 of consensus"  
 2699..3183  
 /note="MER74A repeat: matches 5..528 of consensus"  
 3218..3373  
 /note="L2 repeat: matches 2289..2474 of consensus"  
 4524..4982  
 /note="MTRIC repeat: matches 1..466 of consensus"  
 5006..5377  
 /note="MTR1A2 repeat: matches 1..374 of consensus"  
 5614..6265  
 /note="L1M4 repeat: matches 3331..3985 of consensus"  
 complement(6016..6421)  
 /note="match: SRS: Em:G53026"  
 6368..6618  
 /note="MTR1F repeat: matches 1..255 of consensus"  
 6619..7146  
 /note="MER1A repeat: matches 2..527 of consensus"  
 7147..7238  
 /note="MTR1F repeat: matches 255..347 of consensus"  
 7239..7538  
 /note="AluY repeat: matches 1..302 of consensus"  
 7539..7674  
 repeat\_region

RESULT 8  
 AL354809 135619 bp DNA linear PRI 25-FEB-2001  
 LOCUS Human DNA sequence from clone RP11-527F15 on chromosome 13,  
 complete sequence.  
 ACCESSION AL354809  
 VERSION AL354809.12 GI:13161606  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 135619)  
 AUTHORS Tromans,A.

```
/note="MLTIF repeat: matches 347. .472 of consensus"
repeat_region 11850. 12435
/note="MER21B repeat: matches 100. .660 of consensus"
repeat_region 12624. 12694
/note="MER21B repeat: matches 716. .787 of consensus"
repeat_region 12948. 13260
/note="L1M8 repeat: matches 5623. .5933 of consensus"
repeat_region 14557. 14848
/note="L1R16A repeat: matches 7. .411 of consensus"
repeat_region 16437. 16488
/note="L2 repeat: matches 2643. .2694 of consensus"
repeat_region 17820. 18331
/note="MLTIF repeat: matches 1. .541 of consensus"
repeat_region 18332. 18430
/note="L2 repeat: matches 2641. .2744 of consensus"
repeat_region 19075. 19945
/note="L1M4 repeat: matches 4437. .5313 of consensus"
repeat_region 20939. 21003
/note="MIR repeat: matches 84. .144 of consensus"
repeat_region 22102. 22624
/note="L1M3 repeat: matches 7189. .7739 of consensus"
repeat_region 22625. 22799
/note="AluJ repeat: matches 122. .303 of consensus"
repeat_region 22834. 22968
/note="AluJ repeat: matches 1. .133 of consensus"
repeat_region 22973. 23766
/note="L1M3 repeat: matches 6414. .7196 of consensus"
repeat_region 23771. 24574
/note="L1M3 repeat: matches 5433. .6288 of consensus"
repeat_region 24855. 24958
/note="MIR repeat: matches 112. .215 of consensus"
repeat_region 26923. 26986
/note="32 copies 2 mer ag 90% conserved"
repeat_region 27043. 27129
/note="L2 repeat: matches 2432. .2520 of consensus"
repeat_region 27153. 27242
/note="L2 repeat: matches 2620. .2710 of consensus"
repeat_region 28309. 28610
/note="AluSg repeat: matches 1. .303 of consensus"
repeat_region 31533. 31624
/note="L1P3 repeat: matches 5. .96 of consensus"
repeat_region 31623. 34630
/note="L1P repeat: matches 847. .3842 of consensus"
repeat_region 34722. 34930
/note="L2 repeat: matches 2192. .2404 of consensus"
repeat_region 36155. 36927
/note="MER21B repeat: matches 8. .790 of consensus"
repeat_region 37387. 37909
/note="L2 repeat: matches 2158. .2708 of consensus"
repeat_region 39234. 39522
/note="MLT2E repeat: matches 118. .397 of consensus"
repeat_region 39853. 40305
/note="MLT2B repeat: matches 1. .444 of consensus"
repeat_region 41902. 41949
/note="24 copies 2 mer aa 72% conserved"
repeat_region 42011. 42058
/note="MLT2B repeat: matches 367. .414 of consensus"
repeat_region 42059. 42086
/note="7 copies 4 mer g1gt 100% conserved"
repeat_region 42089. 42452
/note="MLT2B repeat: matches 8. .374 of consensus"
repeat_region 42831. 43032
/note="TIGER2 repeat: matches 6. .218 of consensus"
repeat_region 43033. 43256
/note="AluJb repeat: matches 1. .225 of consensus"
repeat_region 43257. 43521
/note="TIGER2 repeat: matches 218. .446 of consensus"
repeat_region 43525. 43681
/note="AluJb repeat: matches 147. .311 of consensus"
repeat_region 43683. 43892
/note="TIGER2 repeat: matches 530. .759 of consensus"
repeat_region 44049. 44120
/note="AluJ repeat: matches 223. .296 of consensus"

repeat_region 44121. 44271
/note="TIGER2 repeat: matches 2552. .2708 of consensus"
repeat_region 45039. 45252
/note="L1P8 repeat: matches 5950. .6163 of consensus"
repeat_region 45253. 45309
/note="L1P3 repeat: matches 5681. .5737 of consensus"
repeat_region 45316. 45444
/note="L2 repeat: matches 2334. .2462 of consensus"
repeat_region 46424. 46491
/note="L2 repeat: matches 2581. .2652 of consensus"
repeat_region 46534. 46620
/note="L2 repeat: matches 2586. .2677 of consensus"
repeat_region 47570. 47684
/note="FLAM_C repeat: matches 1. .117 of consensus"
repeat_region 47698. 50321
/note="L1P11 repeat: matches 3529. .6165 of consensus"
repeat_region 50339. 50386
/note="24 copies 2 mer aa 83% conserved"
repeat_region 50389. 50659
/note="AluJo repeat: matches 1. .282 of consensus"
repeat_region 50660. 50777
/note="59 copies 2 mer ta 75% conserved"
repeat_region 50662. 50777
/note="29 copies 4 mer tata 75% conserved"
repeat_region 51545. 51618
/note="L1M3 repeat: matches 5883. .5952 of consensus"
repeat_region 51619. 51979
/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region 51980. 52081
/note="L1M3 repeat: matches 5804. .5883 of consensus"
repeat_region 52121. 52465
/note="L1M3 repeat: matches 5374. .5715 of consensus"
repeat_region 52466. 52624
/note="FAM repeat: matches 1. .165 of consensus"
repeat_region 52625. 52945
/note="L1M3 repeat: matches 5060. .5374 of consensus"
repeat_region 53108. 53409
/note="AluSg repeat: matches 1. .301 of consensus"
repeat_region 53907. 54022
/note="L1M repeat: matches 3305. .5433 of consensus"
repeat_region 55475. 55613
/note="MIR repeat: matches 3. .146 of consensus"

Query Match 61.3% Score 508.4; DB 9; Length 135619;
Best Local Similarity 95.8%; Pred. No. 2,1e-133;
Matches 344; Conservative 0; Mismatches 21; Indels 3; Gaps 2;
```

```

Db      102385 CCCGAGCTCTGTAGAGAAATGCAATCTGACTCTCAAGATGTAGAGACACAGCTTCA 102444
Oy      626 TTTCGTGTGGGGAGACACACGACACCTGCTGCCGAGCTTGAATC--TTTTTG 683
Db      102445 TTTCGTGTATCTCAAGACACGACACACCTGCTGCCGAGCTTGAATC--TTTTTG 102504
Oy      684 TTTCGTCAAGTTCTAGAGAGCCCTTCCGTAATATATCTGTTGTCATAGTTCTCTTT 743
Db      102505 TTGCTCAAGTTCTAGAGAGCCCTTCCGTAATATATCTGTTGTCATAGTTCTCTTT 102564
Oy      744 CAAAGTAGTAACCTTTCTATTCTTCTA 771
Db      102565 CAAAGTAGTAACCTTTCTATTCTTCTA 102592

RESULT 9
LOCUS   AL513495                      147310 bp    DNA          linear   HTG 13-JUN-2001
DEFINITION Homo sapiens chromosome 13 clone RP11-466M19, *** SEQUENCING IN
ACCESSION AL513495.1 GI:12733863
VERSION   AL513495.1 GI:12733863
KEYWORDS  HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 147310)
AUTHORS   Burton,J.
TITLE     Direct Submission
JOURNAL   Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
           CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone.
           requests: clonerequests@sanger.ac.uk
           ----- Genome Center
           Center: Sanger Centre
           Center code: SC
           Web site: http://www.sanger.ac.uk
           Contact: humquerry@sanger.ac.uk
           ----- Project Information
           Center project name: ba466M19
           ----- Summary Statistics
           Sequencing vector: plasmid; L08752; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Consensus quality: 138138 bases at least Q40
           Consensus quality: 141205 bases at least Q30
           Consensus quality: 143097 bases at least Q20
           Insert size: 145810; sum-of-contrigs
           Insert size: 162901; agarose-fp
           Quality coverage: 3.13x in Q20 bases; sum-of-contrigs Quality
           coverage: 2.95x in Q20 bases; agarose-fp
           -----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 16 contrigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contrigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *
           * 1
           * 3020 3119: contrig of 3019 bp in length
           * 3120 12625: contrig of 9506 bp in length
           * 12626 12725: gap of 100 bp
           * 12726 20101: contrig of 7376 bp in length
           * 20102 20201: gap of 100 bp
           * 20202 23460: contrig of 3259 bp in length
           * 23461 23560: gap of 100 bp
           * 23561 34856: contrig of 11296 bp in length
           * 34857 34956: gap of 100 bp
           * 34957 50993: contrig of 16037 bp in length
           * 50994 51093: gap of 100 bp
           * 51094 62025: contrig of 10932 bp in length

```

```

FEATURES
source
1..147310
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-466M19"
/clone_1fb="RPC1-11.2"
1..3019
/note="assembly-fragment:00352
fragment_chain:1"
3120..12625
/note="assembly-fragment:01013
fragment_chain:1"
12726..20101
/note="assembly-fragment:00687
fragment_chain:1"
20202..23460
/note="assembly-fragment:00760
fragment_chain:1"
23561..34856
/note="assembly-fragment:00179
fragment_chain:1"
34957..50993
/note="assembly-fragment:00160
fragment_chain:2"
51094..62025
/note="assembly-fragment:00712
fragment_chain:2"
62126..65299
/note="assembly-fragment:00413
fragment_chain:3"
65400..74814
/note="assembly-fragment:00763
fragment_chain:3"
74915..78856
/note="assembly-fragment:00035"
78957..106521
/note="assembly-fragment:00290"
106622..119811
/note="assembly-fragment:00296"
119912..124426
/note="assembly-fragment:00410"
124527..137019
/note="assembly-fragment:00478"
137120..141832
/note="assembly-fragment:00564"
141933..147310
/note="assembly-fragment:00307
vector_side:right"
BASE COUNT 44700 a 28018 c 28143 g 44941 t 1508 others
Query Match 61.3%; Score 508.4; DB 2; Length 147310;
Best Local Similarity 95.8%; Pred. No. 2.1e-133;

```

Matches	544:	Conservative	0:	Mismatches	21:	Indels	3:	Gaps	2:
OY	207	GTTCGTCGTCCCGCCTGTACATGAGGACACACCACTCGTCTGGAATGGCAATGCTGT							266
		TT							
Db	51966	GTGCACACTGTCGCCGCTGTACATGAGGACACACCACTCGTCTGGAATGGCAATGCTGT							52025
OY	267	CAGGACAAAGAAATCCTGTAGTGAATGTTTGAATGTGCTTCACAGAGTTCCAAATGA							326
Db	52026	CAGGACAAAGAAATCCTGTAGTGAATGTTTGAATGTGCTTCACAGAGTTCCAAATGA							52085
OY	327	GCCTGTAGAGCTGCCACCCCTGTTCATGATGTAACACACCAACAGACAGACAGTCTCTG							386
Db	52086	GCCTGTAGAGCTGCCACCCCTGTTCATGATGTAACACCAACAGACAGACAGTCTCTG							52145
OY	387	TTGTCAATCTGTGATCATAGTAAGATTGG-AGGGACAAACACACGGGACTTCAACGAAAC							445
Db	52146	TTGTGAATCTGTGATCATAGTAAGATTGGAAAGGGGAAACAAAGGGGACTTCAACGAAAC							52205
OY	446	TTCAATCTGTGACCGGCCAGGCGCTCAACCAGCAACAGATGTGGAAATGCAAGTGAATCAG							505
Db	52206	TTCAATCTGTGACCGGCCAGGCGCTCAACCAGCAACAGATGTGGAAATGCAAGTGAATCAG							52265
OY	506	TTCCGCTTCCAGAGACTGGCCAGCTATGATGGGGGTGGCAGAGATCTCTTGTCTTCAATCAG							565
Db	52266	TTCCGCTTCCAGAGACTGGCCAGCTATGATGGGGGTGGCAGAGATCTCTTGTCTTCAATCAG							52325
OY	566	CCCTAGCTCTGTAGAGAAATGCAAACTCGACTCAAGATGTGAGAAACACAACTTCA							625
Db	52326	CCCGAGCTCTGTAGAGAAATGCAAACTCGACTCAAGATGTGAGAAACACAACTTCA							52385
OY	626	TTTCTGTGTTGGGGAACACTGACACTGACCTGAGCCAGGTTTAATCTC--TTT							683
Db	52386	TTTCTGTGTTGGGGAACACTGACACTGACCTGAGGAGGTTTAATCTC							52445
OY	684	TTGCTCAAGTCTGTAGAGATCCCTTTCCTGATATATATCTGTTGTCAATGTTCCCTTT							743
Db	52446	TTGCTCAAGTCTGTAGAGATCCCTTTCCTGATATATATCTGTTGTCAATGTTCCCTTT							52505
OY	744	CAAAAGTAGTAACCTTTTCTATTTTCTCA							771
Db	52506	CAAAAGTAGTAACCTTTTCTATTTTCTCA							52533
RESULT 10									
LOCUS	AC110699/c	186415 bp	DNA	linear	HTG 17-JUL-2002				
DEFINITION	Rattus norvegicus clone CH230-157L12, *** SEQUENCING IN PROGRESS								
ACCESSION	AC110699								
VERSION	AC110699.3	GI:21744668							
KEYWORDS	HTG; HTGS_PHASE1.								
SOURCE	Norway rat.								
ORGANISM	Rattus norvegicus								
REFERENCE	1 (bases 1 to 186415)								
AUTHORS	Munhy,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amaralune,H.C., Are,J.R., Ayelo,M., Banks,T., Barbado,J., Benton,J., Bimaje,K., Blankenburg,K., Bonnin,D., Bouck,J., Boyle,S., Brieval,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C., Cox,C., Coyle,M.D., Daborne,S.R., David,R., Davila,M.L., Davis,C., Davy-carroll,L., Dederlich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,								

Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
 Homsl, F., Howard, S., Huber, J., Huyl, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Joliver, S., Joudah, S.,  
 Katslson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kover, C.,  
 Kravtovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtenarge, O., Liu, C., Liu, J., Liu, M., Lounsged, H.,  
 Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mel, G., Mettler, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokkenko, S., Ogun, M., Okunodu, G.,  
 Ordunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Qulles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojoubkan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Slason, I.,  
 Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Swalek, A., Tabori, P., Tamerlis, A., Tamerlis, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, O.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstock, G. and Gbbs, R.  
 -----  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 186415)  
 Worley, K.C.  
 -----  
 Direct Submission  
 Submitted (15-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 186415)  
 Worley, K.C.  
 -----  
 Direct Submission  
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 13, 2002 this sequence version replaced g1:18767328.  
 -----  
 Genome Center  
 -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc.help@bcm.tmc.edu  
 -----  
 Project Information  
 Center project name: GSKZ  
 Center clone name: CH230-157L12  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye, 100% of reads  
 Assembly program: Phrap, version 0.990329  
 Consensus quality: 139868 bases at least Q40  
 Consensus quality: 145668 bases at least Q30  
 Consensus quality: 151388 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html))  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 54 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 1280: contig of 1280 bp in length  
 \* 1281 1380: gap of unknown length  
 \* 1381 2779: contig of 139 bp in length  
 \* 2780 2880: gap of unknown length  
 \* 2880 4438: contig of 1559 bp in length  
 \* 4439 4539: gap of unknown length  
 \* 4539 6012: contig of 1474 bp in length  
 \* 6012 6112: gap of unknown length

6113 7635: contig of 1523 bp in length  
7636 7735: gap of unknown length  
7736 9209: contig of 1474 bp in length  
9210 9309: gap of unknown length  
10427 10427: contig of 1118 bp in length  
10428 10527: gap of unknown length  
10528 11603: contig of 1076 bp in length  
11604 11703: gap of unknown length  
11704 13101: contig of 1398 bp in length  
13102 13201: gap of unknown length  
13202 14668: contig of 1467 bp in length  
14669 14768: gap of unknown length  
14769 15842: contig of 1074 bp in length  
15843 15942: gap of unknown length  
15943 17269: contig of 1327 bp in length  
17270 17369: gap of unknown length  
17370 18936: contig of 1567 bp in length  
18937 19036: gap of unknown length  
19037 20110: contig of 1074 bp in length  
20111 20210: gap of unknown length  
20211 22223: contig of 2013 bp in length  
22224 22323: gap of unknown length  
22324 23473: contig of 1150 bp in length  
23474 23573: gap of unknown length  
23574 24899: contig of 1326 bp in length  
24900 24999: gap of unknown length  
25000 26247: contig of 1248 bp in length  
26248 26347: gap of unknown length  
26348 27995: contig of 1648 bp in length  
27996 28095: gap of unknown length  
28096 29735: contig of 1640 bp in length  
29736 29835: gap of unknown length  
29836 31035: contig of 1200 bp in length  
31036 31135: gap of unknown length  
31136 32181: contig of 1046 bp in length  
32182 32281: gap of unknown length  
32282 34586: contig of 2305 bp in length  
34587 34686: gap of unknown length  
34687 35875: contig of 1189 bp in length  
35876 35975: gap of unknown length  
35976 37133: contig of 1158 bp in length  
37134 37233: gap of unknown length  
37234 39149: contig of 1916 bp in length  
39150 39249: gap of unknown length  
39250 41559: contig of 2310 bp in length  
41560 41659: gap of unknown length  
41660 43823: contig of 2164 bp in length  
43824 43923: gap of unknown length  
43924 46913: contig of 2990 bp in length  
46914 47013: gap of unknown length  
47014 49542: contig of 2529 bp in length  
49543 49642: gap of unknown length  
49643 51523: contig of 1881 bp in length  
51524 51623: gap of unknown length  
51624 53986: contig of 2363 bp in length  
53987 54086: gap of unknown length  
54087 57318: contig of 3232 bp in length  
57319 57418: gap of unknown length  
57420 60601: contig of 3183 bp in length  
60602 60701: gap of unknown length  
60702 63800: contig of 3099 bp in length  
63801 63900: gap of unknown length  
63901 67778: contig of 3878 bp in length  
67779 67878: gap of unknown length  
67879 70679: contig of 2801 bp in length  
70680 70779: gap of unknown length  
70780 73924: contig of 3145 bp in length  
73925 74024: gap of unknown length  
74025 77919: contig of 3895 bp in length  
77920 78019: gap of unknown length  
78020 83314: contig of 5295 bp in length  
83315 83414: gap of unknown length  
83415 87408: contig of 3994 bp in length

87409 87508: gap of unknown length  
87509 91364: contig of 3856 bp in length  
91365 91464: gap of unknown length  
91465 95033: contig of 3569 bp in length  
95034 95133: gap of unknown length  
95134 98981: contig of 3848 bp in length  
98982 99081: gap of unknown length  
99082 104913: contig of 5832 bp in length  
104914 105013: gap of unknown length  
105014 110056: contig of 5043 bp in length  
110057 110156: gap of unknown length  
110157 116497: contig of 6341 bp in length  
116498 116597: gap of unknown length  
116598 121911: contig of 5314 bp in length  
121912 122011: gap of unknown length  
122012 127568: contig of 5557 bp in length  
127569 127668: gap of unknown length  
127669 134565: contig of 6897 bp in length  
134566 134665: gap of unknown length  
134666 145980: contig of 11315 bp in length  
145981 146080: gap of unknown length  
146081 154722: contig of 8642 bp in length  
154723 154822: gap of unknown length  
154823 166077: contig of 11255 bp in length  
166078 166177: gap of unknown length  
166178 186415: contig of 20238 bp in length.

Query Match 59.3%; Score 492; DB 2; Length 186415;  
Best Local Similarity 80.6%; Pred. No. 1e-128;  
Matches 651; Conservative 0; Mismatches 120; Indels 37; Gaps 5;

QY 6 TTGGCATCAGCTGCTCTCTCAACCTTCTCCGACGCCCTGTTCCCAAG 65  
DB 161833 TTGGATTCATCTTGCTTACCTTCTGCTCCACAGTCTGATCCCAAG 161774  
QY 66 CAGAGAAATACCTGCTGAGACCTCTCTCATAGAACAGAGATGATGAT 125  
DB 161773 AGCGCGCAT---CTGGCAGAGGCGCTGCGACAGACCAGAGATGATGAT 161717  
QY 126 TCAAGACCTATGATGATGATGATGATGATGATGATGATGATGATGAT 185  
DB 161716 TCAAGACCTATGATGATGATGATGATGATGATGATGATGATGATGAT 161657  
QY 186 CCACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 245  
DB 161656 CTACATGATGATGATGATGATGATGATGATGATGATGATGAT 161597  
QY 246 TCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 305  
DB 161596 TATGATGATGATGATGATGATGATGATGATGATGATGAT 161537  
QY 306 CTTCCAGCAGTTCATGATGATGATGATGATGATGATGATGATGAT 385  
DB 161536 CTTCCAGTATGATGATGATGATGATGATGATGATGATGATGAT 161477  
QY 366 CAAAGCAGACGAGGCTGTTGATGATGATGATGATGATGATGATGAT 425  
DB 161476 CAAAGCAGACGAGGCTGTTGATGATGATGATGATGATGATGAT 161417  
QY 426 AACGGACTTCAACAGAACTTCAACGCGCCAGCCAGCCAGCAACAGAT 485  
DB 161416 AACGGACTTCAACAGAACTTCAACGCGCCAGCCAGCCAGCAACAGAT 161357  
QY 486 GGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 545  
DB 161356 GGAAGATGATGATGATGATGATGATGATGATGATGATGAT 161297  
QY 546 GGTCTCTTGTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 605  
DB 161296 GGTCTCTTGTCTTCACTTCACTTCACTTCACTTCACTTCACT 161243  
QY 606 ATGTGAGAACCAAGTCTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 665  
DB 161243 ATGTGAGAACCAAGTCTTCTGTTGTTGTTGTTGTTGTTGTTGTT 161184

[illegible]

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W., Lou, S., Lou, S.,  
Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Marpa, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhney, E., McLeod, M., Meador, M., Mei, G., Metzger, N.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, N., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogun, M., Okwuonu, G.,  
Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Scherer, S., Scott, G., Shen, H., Shoostari, N., Slason, I.,  
Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Swatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Tellrod, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 182118)  
Worley, K.C.  
Submitted (19-JUN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center of Medicine -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information -----  
Center project name: GYDN  
Center clone name: CH230-326L8  
----- Summary Statistics -----  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 118171 bases at least Q40  
Consensus quality: 125368 bases at least Q30  
Consensus quality: 130285 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 69 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1  
1004: contig of 1004 bp in length  
1005  
1105: gap of unknown length  
1105  
2235: contig of 1131 bp in length  
2236  
2336: gap of unknown length  
2336  
3480: contig of 1145 bp in length  
3481  
3580: gap of unknown length  
3581  
4634: contig of 1054 bp in length  
4635  
4734: gap of unknown length  
4735  
5901: contig of 1167 bp in length  
5902  
6001: gap of unknown length  
6002  
7042: contig of 1041 bp in length  
7043  
7142: gap of unknown length  
7143  
8354: contig of 1212 bp in length  
8355  
8455: gap of unknown length  
8456  
9464: contig of 1010 bp in length  
9465  
9564: gap of unknown length  
9565  
10594: contig of 1030 bp in length  
10595  
10694: gap of unknown length  
10695  
12121: contig of 1427 bp in length  
12122  
12221: gap of unknown length

12222  
13564: contig of 1343 bp in length  
13565  
13664: gap of unknown length  
13665  
14776: contig of 1112 bp in length  
14777  
14876: gap of unknown length  
14877  
16124: contig of 1248 bp in length  
16125  
16224: gap of unknown length  
16225  
17356: contig of 1132 bp in length  
17357  
17456: gap of unknown length  
17457  
18699: contig of 1243 bp in length  
18700  
18799: gap of unknown length  
18800  
20696: contig of 1897 bp in length  
20697  
20797: gap of unknown length  
20798  
21843: contig of 1047 bp in length  
21844  
21943: gap of unknown length  
21944  
23254: contig of 1311 bp in length  
23255  
23354: gap of unknown length  
23355  
24597: contig of 1243 bp in length  
24598  
24697: gap of unknown length  
24698  
26076: contig of 1379 bp in length  
26077  
26176: gap of unknown length  
26177  
27645: contig of 1469 bp in length  
27646  
27745: gap of unknown length  
27746  
29460: contig of 1715 bp in length  
29461  
29560: gap of unknown length  
29561  
30694: contig of 1134 bp in length  
30695  
30794: gap of unknown length  
30795  
32923: contig of 2129 bp in length  
32924  
33023: gap of unknown length  
33024  
34861: contig of 1838 bp in length  
34862  
34961: gap of unknown length  
34962  
37952: contig of 2991 bp in length  
37953  
38052: gap of unknown length  
38053  
40327: contig of 2175 bp in length  
40328  
40327: gap of unknown length  
40328  
42768: contig of 2441 bp in length  
42769  
42768: gap of unknown length  
42769  
44526: contig of 1658 bp in length  
44527  
44626: gap of unknown length  
44627  
46017: contig of 1391 bp in length  
46018  
46117: gap of unknown length  
46118  
48882: contig of 2765 bp in length  
48883  
48982: gap of unknown length  
48984  
50447: contig of 1465 bp in length  
50448  
50547: gap of unknown length  
50549  
52563: contig of 2016 bp in length  
52564  
52663: gap of unknown length  
52664  
53812: contig of 1149 bp in length  
53813  
53912: gap of unknown length  
53913  
55339: contig of 1337 bp in length  
55340  
55339: gap of unknown length  
55340  
57051: contig of 1712 bp in length  
57052  
57151: gap of unknown length  
57152  
58434: contig of 1283 bp in length  
58435  
58534: gap of unknown length  
58535  
59907: contig of 1373 bp in length  
59908  
60007: gap of unknown length  
60008  
62417: contig of 2410 bp in length  
62418  
62517: gap of unknown length  
62519  
65195: contig of 2678 bp in length  
65196  
65295: gap of unknown length  
65296  
67509: contig of 2214 bp in length  
67510  
67609: gap of unknown length  
67610  
69530: contig of 1921 bp in length  
69531  
69630: gap of unknown length  
69632  
71723: contig of 2093 bp in length  
71724  
71823: gap of unknown length  
71824  
74364: contig of 2441 bp in length  
74365  
74364: gap of unknown length  
74365  
76765: contig of 2401 bp in length  
76766  
76865: gap of unknown length  
76866  
79018: contig of 2153 bp in length  
79019  
79118: gap of unknown length  
82569: contig of 3451 bp in length



82570 82669: gap of unknown length  
 \* 82670 84299: contig of 1630 bp in length  
 \* 84300 84399: gap of unknown length  
 \* 84400 87075: contig of 2676 bp in length  
 \* 87175 87175: gap of unknown length  
 \* 87176 91309: contig of 4134 bp in length  
 \* 91310 91409: gap of unknown length  
 \* 91410 93481: contig of 2072 bp in length  
 \* 93482 93482: gap of unknown length  
 \* 93582 97219: contig of 3638 bp in length  
 \* 97220 97319: gap of unknown length  
 \* 97320 100038: contig of 2719 bp in length  
 \* 100039 100138: gap of unknown length  
 \* 100139 103847: contig of 3709 bp in length  
 \* 103848 103948: gap of unknown length  
 \* 103948 107670: contig of 3723 bp in length  
 \* 107671 107770: gap of unknown length  
 \* 107771 110595: contig of 2825 bp in length  
 \* 110596 110696: gap of unknown length  
 \* 110696 115142: contig of 4447 bp in length  
 \* 115143 115143: gap of unknown length

Query Match 43.1%; Score 357.8; DB 2; Length 182118;  
 Best Local Similarity 81.2%; Pred. No. 1.8e-90;  
 Matches 478; Conservative 0; Mismatches 77; Indels 34; Gaps 4;

225 ACATGGGACACGACCCCTGCTGGAATGCAATGCTGTTTCAGAGACAAATCTTGA 284  
 Db ACATGGGACACGACCCCTGCTGGAATGCAATGCTGTTTCAGAGACAAATCTTGA 32838  
 285 GTGAGTTTGTGAATGTTGCTTCCAGCGAGTTCACAAATCAGCGTGTAGACCTCCAGC 344  
 Db GTGAGTTTGTGAATGTTGCTTCCAGCGAGTTCACAAATCAGCGTGTAGACCTCCAGC 32778  
 32837 GTGAGTTTGTGAATGTTGCTTCCAGCGAGTTCACAAATCAGCGTGTAGACCTCCAGC 32778  
 345 CTGTTTCATGATGAGACCCACACAGCCAGCAGCGTCTTGTGTCATCTGTGGATGAG 404  
 Db CTGTTTCATGATGAGACCCACACAGCCAGCAGCGTCTTGTGTCATCTGTGGATGAG 32718  
 32777 CTGTTTCATGATGAGACCCACACAGCCAGCAGCGTCTTGTGTCATCTGTGGATGAG 32718  
 405 TGAAGTTTGAAGGAGAACAAACAGCGGACTTCACAGAACTTCTGACCGCCAGC 464  
 Db TGAAGTTTGAAGGAGAACAAACAGCGGACTTCACAGAACTTCTGACCGCCAGC 32658  
 32717 TGAAGTTTGAAGGAGAACAAACAGCGGACTTCACAGAACTTCTGACCGCCAGC 32658  
 465 CCTCACCACAGAACACACTGTGGAAGATGCAAGTGTGCTCCGCTCCAGGACTGG 524  
 Db CCTCACCACAGAACACACTGTGGAAGATGCAAGTGTGCTCCGCTCCAGGACTGG 32598  
 32657 CGTACACCAGAACACACTGTGGAAGATGCAAGTGTGCTCCGCTCCAGGACTGG 32598  
 525 CCAGCTAGTGGGGGAGGAGAGTCTTTGCTTATTCAGCCCTAGCTCTAGAGAAA 584  
 Db CCAGCTAGTGGGGGAGGAGAGTCTTTGCTTATTCAGCCCTAGCTCTAGAGAAA 32543  
 32597 CCAGCTAGTGGGGGAGGAGAGTCTTTGCTTATTCAGCCCTAGCTCTAGAGAAA 32543  
 585 TGCACACCTCGACTCTCAAGTGTGAGGAACACAGTTCATTTCTGTTGTCGAGAGC 644  
 Db TGCACACCTCGACTCTCAAGTGTGAGGAACACAGTTCATTTCTGTTGTCGAGAGC 32484  
 32542 TGCCAGTCTCAAAATCTC-AGGACGTGAGAGAACACTAGTTTGTGTCGAGAGC 32484  
 645 ACTGAGACTCCACTGTGCGGAGGTGAACCTTTTGTGTCCTCAAGTTCAGAGAGTCC 704  
 Db ACTGAGACTCCACTGTGCGGAGGTGAACCTTTTGTGTCCTCAAGTTCAGAGAGTCC 32451  
 32483 ACTGAGACTCCACTGTGCGGAGGTGAACCTTTTGTGTCCTCAAGTTCAGAGAGTCC 32451  
 705 CTTCCTGAATATATATCTGTTGTCATAGTTCTCTTTCACAAAGTAGTAACTTTCTAT 764  
 Db CTTCCTGAATATATATCTGTTGTCATAGTTCTCTTTCACAAAGTAGTAACTTTCTAT 32450  
 32450 CTTCCTGAATATATATCTGTTGTCATAGTTCTCTTTCACAAAGTAGTAACTTTCTAT 32392  
 765 TTTTCTACTGCGCAGTAGAGACTGTATCTGGAATTCGACAAATA 813  
 Db TTTTCTACTGCGCAGTAGAGACTGTATCTGGAATTCGACAAATA 32343  
 32391 TTTTCTACTGCGCAGTAGAGACTGTATCTGGAATTCGACAAATA 32343

RESULT 13  
 AX261550 318 bp DNA linear PAT 26-OCT-2001  
 LOCUS AX261550  
 DEFINITION Sequence 1201 from Patent WO0173027.  
 ACCESSION AX261550  
 VERSION AX261550.1 GI:15510517

KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 Meagher, M.J., Xu, J. and King, G.E.  
 TITLES  
 JOURNAL Composition and methods for therapy and diagnosis of colon cancer  
 PATENT: WO 0173027-A 1201 04-OCT-2001;  
 CORNIX CORPORATION (US)  
 FEATURES  
 SOURCE Location/Qualifiers  
 1. 318  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 76 a 82 c 83 g 76 t 1 others  
 ORIGIN

Query Match 38.2%; Score 317; DB 6; Length 318;  
 Best Local Similarity 99.7%; Pred. No. 4.3e-79;  
 Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

147 CCTGACAGCTGCTGAGAGATTTTGTGATGCTCTACTACACCATGATAGCGGCGC 206  
 Db CCTGACAGCTGCTGAGAGATTTTGTGATGCTCTACTACACCATGATAGCGGCGC 60  
 1 CCTGACAGCTGCTGAGAGATTTTGTGATGCTCTACTACACCATGATAGCGGCGC 60  
 207 GTTTCCTGCTCCGCGCTGTACATGAGGACACGACACCCCTGCTGGAATGCAATGCTGTT 266  
 Db GTTTCCTGCTCCGCGCTGTACATGAGGACACGACACCCCTGCTGGAATGCAATGCTGTT 120  
 61 GTTTCCTGCTCCGCGCTGTACATGAGGACACGACACCCCTGCTGGAATGCAATGCTGTT 120  
 267 CAGGACAAAGATTCCTTGAAGTGTGTTTGAATGTTGCTTCCAGGAGTTCACAAATCA 326  
 Db CAGGACAAAGATTCCTTGAAGTGTGTTTGAATGTTGCTTCCAGGAGTTCACAAATCA 180  
 121 CAGGACAAAGATTCCTTGAAGTGTGTTTGAATGTTGCTTCCAGGAGTTCACAAATCA 180  
 327 GCGTGTGATGCTGCTGAGCTGCTGTTGATGATGAGAACCCACACAGCCAGCGTCTTGG 386  
 Db GCGTGTGATGCTGCTGAGCTGCTGTTGATGATGAGAACCCACACAGCCAGCGTCTTGG 240  
 181 GCGTGTGATGCTGCTGAGCTGCTGTTGATGATGAGAACCCACACAGCCAGCGTCTTGG 240  
 387 TTGTCATCTGTGATGATGAGGAGTGTGAGGAGAACAAACAGGACTTCACAGAACT 446  
 Db TTGTCATCTGTGATGATGAGGAGTGTGAGGAGAACAAACAGGACTTCACAGAACT 300  
 241 TTGTCATCTGTGATGATGAGGAGTGTGAGGAGAACAAACAGGACTTCACAGAACT 300  
 447 TCATCCTGACCGCCAGC 464  
 Db TCATCCTGACCGCCAGC 318

RESULT 14  
 AC111510/c  
 LOCUS AC111510  
 DEFINITION Rattus norvegicus clone CH230-133M15, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 58 unordered pieces:  
 AC111510  
 AC111510.2 GI:21736059  
 HTG: HTGS PHASE1.  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 153483)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., All-ouman, F.R., Allen, C.,  
 Albrooks, S.L., Amaral, H.C., Are, J.R., Ayele, M., Banks, T.,  
 Barbarella, J., Benton, J., Blum, K., Blankenburg, K., Bonin, D.,  
 Bouck, J., Bowles, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, Z., Chowdhury, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinn, H.H.,  
 Dunham, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabist, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gortell, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, R., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogue, M., Hollaway, C., Hollins, B.,  
 Homsl, F., Howard, S., Huber, J., Hui, S., Hume, J., Jackson, L. E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,  
 Li, J., Li, Z., Licharge, O., Lien, C., Liu, J., Liu, W., Louised, H.,  
 Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M.,  
 Moser, M., Neal, D., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Nguyen, N., Nickerson, E., Nwokwenkwo, S., Ogih, M., Okunolu, G.,  
 Ottaguaye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,  
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, Y., Villalón, D., Vinson, R., Wang, O.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,  
 Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Unpublished  
 Direct Submission  
 2 (bases 1 to 153483)  
 Worley, K.C.  
 Submitted (19-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 153483)  
 Worley, K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced g1:18701275.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Project name: GMLT  
 Center project name: CH230-133N15  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye 100% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 88643 bases at least Q40  
 Consensus quality: 93187 bases at least Q30  
 Consensus quality: 97900 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 58 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 1313: contig of 1313 bp in length  
 \* 1314 1413: gap of unknown length  
 \* 3169 3168: contig of 1755 bp in length  
 \* 3269 3268: gap of unknown length  
 4610: contig of 1342 bp in length  
 4611 4710: gap of unknown length  
 \* 4711 6083: contig of 1373 bp in length  
 \* 6084 6184: gap of unknown length  
 \* 6185 7726: contig of 1542 bp in length  
 \* 7727 7826: gap of unknown length  
 \* 7827 8962: contig of 1137 bp in length  
 \* 8963 9062: gap of unknown length  
 \* 9063 10153: contig of 1090 bp in length  
 \* 10154 10253: gap of unknown length  
 \* 10254 11464: contig of 1211 bp in length  
 \* 11465 11564: gap of unknown length  
 \* 11565 13182: contig of 1619 bp in length  
 \* 13183 13282: gap of unknown length  
 \* 13283 14545: contig of 1263 bp in length  
 \* 14546 14646: gap of unknown length  
 \* 14647 15847: contig of 1202 bp in length  
 \* 15848 15947: gap of unknown length  
 \* 15948 17592: contig of 1645 bp in length  
 \* 17593 17692: gap of unknown length  
 \* 17693 18890: contig of 1198 bp in length  
 \* 18891 18990: gap of unknown length  
 \* 18991 20761: contig of 1770 bp in length  
 \* 20762 20860: gap of unknown length  
 \* 20861 22271: contig of 1411 bp in length  
 \* 22272 22371: gap of unknown length  
 \* 22372 23809: contig of 1438 bp in length  
 \* 23810 25341: gap of unknown length  
 \* 25342 25441: contig of 1432 bp in length  
 \* 25442 26863: gap of unknown length  
 \* 26864 26963: contig of 1422 bp in length  
 \* 26964 28125: gap of unknown length  
 \* 28126 28225: contig of 1162 bp in length  
 \* 28226 29789: gap of unknown length  
 \* 29790 29889: contig of 1564 bp in length  
 \* 29890 31582: gap of unknown length  
 \* 31583 31682: contig of 1693 bp in length  
 \* 31683 32773: gap of unknown length  
 \* 32774 32873: contig of 1091 bp in length  
 \* 32874 34244: gap of unknown length  
 \* 34245 34344: contig of 1371 bp in length  
 \* 34345 36071: gap of unknown length  
 \* 36072 36171: contig of 1727 bp in length  
 \* 36172 37383: gap of unknown length  
 \* 37384 37483: contig of 1212 bp in length  
 \* 37484 38840: gap of unknown length  
 \* 38841 38940: contig of 1357 bp in length  
 \* 38941 40471: gap of unknown length  
 \* 40472 40571: contig of 1531 bp in length  
 \* 40572 42670: gap of unknown length  
 \* 42671 42770: contig of 2099 bp in length  
 \* 42771 45260: gap of unknown length  
 \* 45261 45360: contig of 2490 bp in length  
 \* 45361 47344: gap of unknown length  
 \* 47345 49411: contig of 1984 bp in length  
 \* 49412 49511: gap of unknown length  
 \* 49512 51405: contig of 1967 bp in length  
 \* 51406 51505: gap of unknown length  
 \* 51506 53397: contig of 1894 bp in length  
 \* 53398 53497: gap of unknown length  
 \* 53498 56170: contig of 2673 bp in length  
 \* 56171 56270: gap of unknown length  
 \* 56271 59034: contig of 2764 bp in length  
 \* 59035 59134: gap of unknown length  
 \* 59135 60990: contig of 1856 bp in length  
 \* 60991 61090: gap of unknown length  
 \* 61091 62910: contig of 1820 bp in length  
 \* 62911 63010: gap of unknown length  
 \* 63011 65846: contig of 2338 bp in length  
 \* 65847 67736: gap of unknown length  
 \* 67737 67836: contig of 2090 bp in length  
 \* 67837 67936: gap of unknown length

COMMENT

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

*	67837	70060	contig of 2224 bp in length
*	70061	70160	gap of unknown length
*	70161	72709	contig of 2549 bp in length
*	72710	72809	gap of unknown length
*	72810	75211	contig of 2402 bp in length
*	75212	75311	gap of unknown length
*	75312	77837	contig of 2526 bp in length
*	77838	77937	gap of unknown length
*	77938	80801	contig of 2864 bp in length
*	80802	80901	gap of unknown length
*	80902	85672	contig of 4771 bp in length
*	85673	85772	gap of unknown length
*	85773	89247	contig of 3475 bp in length
*	89248	89347	gap of unknown length
*	89348	92449	contig of 3102 bp in length
*	92450	92549	gap of unknown length
*	92550	94576	contig of 2027 bp in length
*	94577	94677	gap of unknown length
*	94677	96771	contig of 4995 bp in length
*	96772	99771	gap of unknown length
*	99772	106874	contig of 7103 bp in length
*	106875	106974	gap of unknown length
*	106975	112478	contig of 5504 bp in length
*	112479	112578	gap of unknown length
*	112579	119343	contig of 6765 bp in length
*	119344	119443	gap of unknown length
*	119444	123855	contig of 4412 bp in length
*	123856	123955	gap of unknown length
*	123956	129007	contig of 5052 bp in length

Query Match	34.7%;	Score 287.6;	DB 2;	Length 153483
Best Local Similarity	67.3%;	Pred. No. 1.8e-70;		
Matches 381;	Conservative	0;	Mismatches 179;	Indels 6;

[illegible]

```

RESULT 15
AF201942
LOCUS AF201942
DEFINITION Homo sapiens DC9 (DC9) mRNA, complete cds.
ACCESSION AF201942
VERSION AF201942.1
KEYWORDS GI:9295187
SOURCE FLI CDNA.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 896)
AUTHORS Li,Y., Peng,Y., Li,N., Gu,W., Han,Z., Fu,G. and Chen,Z.
TITLE Novel genes expressed in human dendritic cell
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 896)
AUTHORS Li,Y., Peng,Y., Li,N., Gu,W., Han,Z., Fu,G. and Chen,Z.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Chinese National Human Genome Center at
Shanghai, 351 Guo Shoujing Road, zhangjiang Hi-Tech Park, Pudong,
Shanghai 201203, P. R. China
FEATURES
Source
1..896
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="dendritic cell"
1..896

```

DB	Query Match	Best Local Similarity	Matches	Conservative	Score	DB 9:	Length	896:
QY	114	CATCTGTGGATTTC	72.4%	0	251.6	DB 9:	896:	
DB	152	CGTCTGTGGATTTC	0	124	Indels	0	Gaps	0
QY	174	ATGTCTACTACAC	0	124	Indels	0	Gaps	0
DB	212	ATATTACTATGAG	0	124	Indels	0	Gaps	0
QY	234	CAGCCACCCTG	0	124	Indels	0	Gaps	0
DB	272	AGGCCACCCTTA	0	124	Indels	0	Gaps	0
QY	294	TGGAATGTGGCT	0	124	Indels	0	Gaps	0
DB	332	TTGACACATGCT	0	124	Indels	0	Gaps	0
QY	354	ATGAAGCCAC	0	124	Indels	0	Gaps	0
DB	392	AGCAAGCACT	0	124	Indels	0	Gaps	0
QY	414	AGGGAAACAA	0	124	Indels	0	Gaps	0
DB	452	ATGGAACAA	0	124	Indels	0	Gaps	0
QY	474	GCAACACAG	0	124	Indels	0	Gaps	0
DB	512	ACAAATCTGT	0	124	Indels	0	Gaps	0

Mon Jun 30 08:51:04 2003

us-09-763-902b-10.rge

Page 18

Oy 534 GGGGTGCGAGGTCCTTTCCTTCATTC 563  
||| ||| ||| ||| ||| ||| |||  
Db 572 AAAGGGCAAAAGTCATTCATTCATTCGTC 601

Search completed: June 28, 2003, 02:38:50  
Job time : 2249.7 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:20:48 ; Search time 39.6938 Seconds

(without alignments)  
5572.555 Million cell updates/sec

Title: US-09-763-902B-10

Perfect score: 1489

Sequence: 1 tgaacttgcatcaccgtg.....ataattaataacacatg 830

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Xgapop 10.0	Xgapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+2.n2p.model -DEV=xlp  
-O=cg22.1/USPRO.spool/US09763902/runat\_24062003\_130340\_12405/app.query.fasta\_1.2254  
-DB=AGeneseq\_101002 -QFWT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR.SCORE=pc -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09763902.ECGN.1.1.162.4runat\_24062003\_130340\_12405 -NCPU=6 -ICPU=3  
-NO.MMAP -LARGESOURCY -NEG.SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: A.Geneseq\_101002:\*

1: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	813	54.6	164	AAU16228
2	732	49.2	140	AAV82318
3	584.5	39.3	197	AAV95379
4	584	39.2	142	AAV82323
5	584	39.2	142	AAV82323
6	579	38.9	162	AAV82327
7	308	20.7	133	AAV82327
8	130	8.7	125	AAV82327
9	120	8.1	153	AAV82327
10	119	8.0	688	AAV82327
11	119	8.0	690	AAV82327
12	117.5	7.9	123	AAV82327
13	117.5	7.9	126	AAV82327
14	117.5	7.9	131	AAV82327
15	117.5	7.9	134	AAV82327
16	115.5	7.8	143	AAV82327
17	105.5	7.1	130	AAV82327
18	105.5	7.1	466	AAV82327
19	105.5	7.1	477	AAV82327
20	105	7.1	122	AAV82327
21	105	7.1	157	AAV82327
22	93	6.2	786	AAV82327
23	91.5	6.1	508	AAV82327
24	91.5	6.1	962	AAV82327
25	90.5	6.1	685	AAV82327
26	90	6.0	466	AAV82327
27	90	6.0	466	AAV82327
28	90	6.0	1061	AAV82327
29	89.5	6.0	659	AAV82327
30	89	6.0	326	AAV82327
31	89	6.0	1609	AAV82327
32	88.5	5.9	129	AAV82327
33	88.5	5.9	130	AAV82327
34	88.5	5.9	218	AAV82327
35	88.5	5.9	671	AAV82327
36	88.5	5.9	671	AAV82327
37	88	5.9	1081	AAV82327
38	88	5.9	1081	AAV82327
39	87.5	5.9	160	AAV82327
40	87.5	5.9	160	AAV82327
41	87.5	5.9	460	AAV82327
42	87.5	5.9	472	AAV82327
43	87	5.8	319	AAV82327
44	87	5.8	364	AAV82327
45	87	5.8	440	AAV82327

## ALIGNMENTS

RESULT 1  
AAU16228  
AAU16228 standard; Protein; 164 AA.

AAU16228; 07-NOV-2001 (first entry)

Human novel secreted protein, Seq ID 1181.

Human; immunosuppressive; antiarthritic; antirheumatic;  
cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
neuroprotective; antibacterial; virucide; fungicide; optalmological;  
vulnerable; secreted protein; rheumatoid arthritis;  
hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
cerebrovascular disorder; cerebral ischemia; angiodysgenesis;  
nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
corneal infection; wound healing; epithelial cell proliferation;  
skin ageing; food additive; preservative; antiproliferative.

XX Homo sapiens.  
 OS  
 XX  
 PN WO20015322-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PE 17-JAN-2001: 2001WO-US01341.  
 XX  
 PR 31-JAN-2000: 2000US-0179065.  
 PR 04-FEB-2000: 2000US-0180628.  
 PR 24-FEB-2000: 2000US-0184664.  
 PR 02-MAR-2000: 2000US-0186350.  
 PR 16-MAR-2000: 2000US-0189874.  
 PR 17-MAR-2000: 2000US-0190076.  
 PR 18-APR-2000: 2000US-0198123.  
 PR 19-MAY-2000: 2000US-0205515.  
 PR 07-JUN-2000: 2000US-0209467.  
 PR 28-JUN-2000: 2000US-0214886.  
 PR 30-JUN-2000: 2000US-0215135.  
 PR 07-JUL-2000: 2000US-0216647.  
 PR 07-JUL-2000: 2000US-0216880.  
 PR 11-JUL-2000: 2000US-0217487.  
 PR 11-JUL-2000: 2000US-0217496.  
 PR 14-JUL-2000: 2000US-0218290.  
 PR 26-JUL-2000: 2000US-0220963.  
 PR 26-JUL-2000: 2000US-0220964.  
 PR 14-AUG-2000: 2000US-0224518.  
 PR 14-AUG-2000: 2000US-0224519.  
 PR 14-AUG-2000: 2000US-0225213.  
 PR 14-AUG-2000: 2000US-0225214.  
 PR 14-AUG-2000: 2000US-0225266.  
 PR 14-AUG-2000: 2000US-0225267.  
 PR 14-AUG-2000: 2000US-0225268.  
 PR 14-AUG-2000: 2000US-0225270.  
 PR 14-AUG-2000: 2000US-0225447.  
 PR 14-AUG-2000: 2000US-0225757.  
 PR 14-AUG-2000: 2000US-0225758.  
 PR 14-AUG-2000: 2000US-0225759.  
 PR 18-AUG-2000: 2000US-0226279.  
 PR 22-AUG-2000: 2000US-0226681.  
 PR 22-AUG-2000: 2000US-0226688.  
 PR 22-AUG-2000: 2000US-0227182.  
 PR 23-AUG-2000: 2000US-0227009.  
 PR 30-AUG-2000: 2000US-0228824.  
 PR 01-SEP-2000: 2000US-0229287.  
 PR 01-SEP-2000: 2000US-0229343.  
 PR 01-SEP-2000: 2000US-0229344.  
 PR 01-SEP-2000: 2000US-0229345.  
 PR 05-SEP-2000: 2000US-0229509.  
 PR 05-SEP-2000: 2000US-0229513.  
 PR 06-SEP-2000: 2000US-0230437.  
 PR 06-SEP-2000: 2000US-0230438.  
 PR 08-SEP-2000: 2000US-0231242.  
 PR 08-SEP-2000: 2000US-0231243.  
 PR 08-SEP-2000: 2000US-0231244.  
 PR 08-SEP-2000: 2000US-0231413.  
 PR 08-SEP-2000: 2000US-0231414.  
 PR 08-SEP-2000: 2000US-0232080.  
 PR 08-SEP-2000: 2000US-0232081.  
 PR 12-SEP-2000: 2000US-0231968.  
 PR 14-SEP-2000: 2000US-0232397.  
 PR 14-SEP-2000: 2000US-0232398.  
 PR 14-SEP-2000: 2000US-0232399.  
 PR 14-SEP-2000: 2000US-0232400.  
 PR 14-SEP-2000: 2000US-0232401.  
 PR 14-SEP-2000: 2000US-0233063.  
 PR 14-SEP-2000: 2000US-0233064.  
 PR 14-SEP-2000: 2000US-0233065.  
 PR 21-SEP-2000: 2000US-0234423.  
 PR 21-SEP-2000: 2000US-0234474.  
 PR 25-SEP-2000: 2000US-0234997.  
 PR 25-SEP-2000: 2000US-0234998.  
 PR 25-SEP-2000: 2000US-0234999.  
 PR 25-SEP-2000: 2000US-0234999.

PR 26-SEP-2000: 2000US-0235484.  
 PR 27-SEP-2000: 2000US-0235834.  
 PR 27-SEP-2000: 2000US-0235836.  
 PR 29-SEP-2000: 2000US-0236327.  
 PR 29-SEP-2000: 2000US-0236367.  
 PR 29-SEP-2000: 2000US-0236367.  
 PR 29-SEP-2000: 2000US-0236368.  
 PR 29-SEP-2000: 2000US-0236369.  
 PR 29-SEP-2000: 2000US-0236370.  
 PR 02-OCT-2000: 2000US-0236802.  
 PR 02-OCT-2000: 2000US-0237037.  
 PR 02-OCT-2000: 2000US-0237038.  
 PR 02-OCT-2000: 2000US-0237039.  
 PR 02-OCT-2000: 2000US-0237040.  
 PR 13-OCT-2000: 2000US-0239937.  
 PR 13-OCT-2000: 2000US-0239937.  
 PR 20-OCT-2000: 2000US-0240960.  
 PR 20-OCT-2000: 2000US-0241221.  
 PR 20-OCT-2000: 2000US-0241785.  
 PR 20-OCT-2000: 2000US-0241786.  
 PR 20-OCT-2000: 2000US-0241787.  
 PR 20-OCT-2000: 2000US-0241808.  
 PR 20-OCT-2000: 2000US-0241809.  
 PR 20-OCT-2000: 2000US-0241826.  
 PR 01-NOV-2000: 2000US-0244617.  
 PR 08-NOV-2000: 2000US-0244617.  
 PR 08-NOV-2000: 2000US-0246474.  
 PR 08-NOV-2000: 2000US-0246475.  
 PR 08-NOV-2000: 2000US-0246476.  
 PR 08-NOV-2000: 2000US-0246477.  
 PR 08-NOV-2000: 2000US-0246478.  
 PR 08-NOV-2000: 2000US-0246523.  
 PR 08-NOV-2000: 2000US-0246524.  
 PR 08-NOV-2000: 2000US-0246525.  
 PR 08-NOV-2000: 2000US-0246526.  
 PR 08-NOV-2000: 2000US-0246527.  
 PR 08-NOV-2000: 2000US-0246528.  
 PR 08-NOV-2000: 2000US-0246532.  
 PR 08-NOV-2000: 2000US-0246609.  
 PR 08-NOV-2000: 2000US-0246610.  
 PR 08-NOV-2000: 2000US-0246611.  
 PR 08-NOV-2000: 2000US-0246613.  
 PR 17-NOV-2000: 2000US-0249207.  
 PR 17-NOV-2000: 2000US-0249208.  
 PR 17-NOV-2000: 2000US-0249209.  
 PR 17-NOV-2000: 2000US-0249210.  
 PR 17-NOV-2000: 2000US-0249211.  
 PR 17-NOV-2000: 2000US-0249212.  
 PR 17-NOV-2000: 2000US-0249213.  
 PR 17-NOV-2000: 2000US-0249214.  
 PR 17-NOV-2000: 2000US-0249215.  
 PR 17-NOV-2000: 2000US-0249216.  
 PR 17-NOV-2000: 2000US-0249217.  
 PR 17-NOV-2000: 2000US-0249218.  
 PR 17-NOV-2000: 2000US-0249244.  
 PR 17-NOV-2000: 2000US-0249245.  
 PR 17-NOV-2000: 2000US-0249264.  
 PR 17-NOV-2000: 2000US-0249265.  
 PR 17-NOV-2000: 2000US-0249266.  
 PR 17-NOV-2000: 2000US-0249267.  
 PR 17-NOV-2000: 2000US-0251179.  
 PR 08-DEC-2000: 2000US-0251856.  
 PR 08-DEC-2000: 2000US-0251868.  
 PR 08-DEC-2000: 2000US-0251869.  
 PR 08-DEC-2000: 2000US-0251899.  
 PR 11-DEC-2000: 2000US-0251990.  
 PR 11-DEC-2000: 2000US-0254097.  
 PR 05-JAN-2001: 2001US-0259678.  
 XX



Sequence 140 AA:  
Alignment Scores:  
Pred. No.: 5,22e-76 Length: 140  
Score: 732.00 Matches: 140  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 49,168 Indels: 0  
DB: 21 Gaps: 0  
US-09-763-902B-10 (1-830) x AAY82318 (1-140)  
QY 110 ATGCGATCGTGGATTTCAAGACATATGATGATGAGCCTCGAGAGCTTGTGAGAGTT 169  
DB 1 MetalaSerValaSPheLysThrYrValaSPGlnAlaCysArgAlaLglnGluPhe 20  
QY 170 GTCATATGCTACTACACCCATGATGATAGCGGCGCTTGTGCTGCTGCTGCTGATGATG 229  
DB 21 ValaSNValaLTYrTThrThrMetaspLysArgArgLeuLeuSerArgLeuTyrMet 40  
QY 230 GGCAGACCCAGCCCTGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 289  
DB 41 GYThrAlaThrLeuValaLrpsnngLysAlaValaSerGlyGlnGlnSerLeuSerGlu 60  
QY 290 TTTTGTGAATGTGCTTCCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 349  
DB 61 PhePheGlnMetLeuProSerSerGluPheGlnLLeSerValaLaspCysGlnProVal 80  
QY 350 CATG 409  
DB 81 HlaSPGlnAlaLThrProSerGlnThrThrValaLLeuValaLLeCysGlySerValLys 100  
QY 410 TTTGAGGGAGAACAAACAGGAGGAGCTTCAACAGAACTTATCTGACGCCAGGAGCTCA 469  
DB 101 PheGlnGlyAlaSNLysGlnLrpsnngLysAlaValaSerGlyGlnGlnSerLeuSerGlu 120  
QY 470 CCCAGACACACAGTGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 529  
DB 121 ProSerAsnThrValaLrpsLysLLeaLaserAspCysPheArgPheGlnAspLrpalaser 140  
RESULT 3  
AAB95379  
ID AAB95379 standard; Protein; 197 AA.  
AC AAB95379;  
DF 26-JUN-2001 (first entry)  
DE Human protein sequence SEQ ID NO:17707.  
KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
OS Homo sapiens.  
PN EP1074617-A2.  
PD 07-FEB-2001.  
PF 28-JUL-2000; 2000EP-0116126.  
PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
PA (HELI-) HELIX RES INST.  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX Claim 8; SEQ ID 17707; 2537pp + CD ROW; English.  
PS The present invention describes primer sets for synthesizing 5602  
XX full-length cDNAs defined in the specification. Where a primer set  
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX  
SQ Sequence 197 AA:  
Alignment Scores:  
Pred. No.: 8,82e-59 Length: 197  
Score: 584.50 Matches: 106  
Percent Similarity: 90.85% Conservative: 23  
Best Local Similarity: 74.65% Mismatches: 12  
Query Match: 39,254 Indels: 1  
DB: 22 Gaps: 1  
US-09-763-902B-10 (1-830) x AAB95379 (1-197)  
QY 107 GAGATGGCA--TCTGTGATTTCAAGACCTATGATGATGATGATGATGATGATGATGATGATGATGATG 163  
DB 55 GlnMetAlaLThrSerLeuAspPheLysThrTyrValaSPGlnAlaCysArgAlaLgln 74  
QY 164 GAGTTGTCAATGTCTACTACACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 223  
DB 75 GluPheValaSNLLeTyrTyrGlnThrMetaspLysArgArgAlaLeuThrArgLeu 94  
QY 224 TGCATGGGACAGCCAGCCCTGGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 283  
DB 95 TyrLeuAspLysAlaThrLeuLLeTyrPasnngLysAlaValaSerGlyLeuAspAlaLeu 114  
QY 284 AGTGAAGTTTGAATGTGCTTCCAGCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 343  
DB 115 AsnAsnPhePheAspThrLeuProSerSerGluPheGlnValaSNMetLeuAspCysGln 134  
QY 344 CCTGTTCATG 403  
DB 135 ProValaLHISGlnGlnAlaLThrGlnSerGlnThrThrValaLeuValaLThrSerGlyThr 154  
QY 404 GTGAAGTTGAGGGAGGACAAACAGCGGACTTCAACAGACTTCTATCTGACCGCCAG 463  
DB 155 ValLysPheAspLysnngLysPhePheAsnGlnAspPheLeuLeuThraLgln 174  
QY 464 GCCTACCCAGACACAGTGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 523  
DB 175 SerThrProAsnAsnThrValaLrpsLysLLeaLaserAspCysPheArgPheGlnAspLrpal 194  
QY 524 GCCAGC 529  
DB 195 SerSer 196



RESULT 4  
AAV82323  
ID AAV82323 standard; Protein: 142 AA.  
XX  
AC AAV82323;  
XX  
DT 19-JUN-2000 (first entry)  
XX  
DE Human protein transport molecule (PRAM) SEQ ID NO:7.  
XX  
KW Human; protein transport molecule; PRAM; diagnosis; cytostatic;  
KW antiarthritic; antidiabetic; immunosuppressant; antiarteriosclerotic;  
KW antiallergic; antidiabetic; antilipemic; antirheumatic; osteopathic;  
KW dermatological; antianaemic; antipruritic; hepatotropic; antispasmodic;  
KW antinflammatory; antihiv; protein transport; regulator; cancer;  
KW immune disorder; cell proliferative disorder; secretory disorder;  
KW urticaria; allergy; abnormal vesicle trafficking; asthma;  
KW autoimmune haemolytic anaemia.  
XX  
OS Homo sapiens.  
XX  
PN WC0200012703-A2.  
XX  
PD 09-MAR-2000.  
XX  
PF 26-AUG-1999; 99MO-US19616.  
XX  
PR 27-AUG-1998; 98US-0098206.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;  
PI Gorgone GA, Baughn MR, Patterson C;  
XX  
DR WPI: 2000-256642/22.  
XX  
DR N-PSDB: AAA08041.  
XX  
PT New human protein transport-associated polypeptide and polynucleotide  
PT useful for diagnosis, prevention and treatment of cell proliferative  
PT and secretory disorders such as leukemia, cystic fibrosis  
XX  
PS Claim 1; Page 68; 75pp; English.  
XX  
CC AAA08035 to AAA08042 encode the human protein transport-associated  
CC molecules (PRAMs) given in AAV82317 to AAV82324. The PRAMs have  
CC cytosolic, antiarthritic, antidiabetic, immunosuppressant,  
CC antiallergic, antidiabetic, antilipemic, antirheumatic, osteopathic,  
CC antirheumatic, osteopathic, dermatological, antianaemic, antipruritic,  
CC hepatotropic, antispasmodic, antinflammatory and antihiv activities, and  
CC regulate protein transport. PRAM proteins and antagonists are useful for  
CC preventing or treating a disorder associated with decreased or increased  
CC expression or activity of PRAM. PRAM polynucleotides are useful for  
CC diagnosing conditions associated with PRAM, comprising detecting PRAM by  
CC forming a hybridisation complex, preferably after PCR amplifying the  
CC biological sample. Diseases prevented, treated or diagnosed include cell  
CC proliferative disorders such as cancers, immune disorders, secretory  
CC disorders and other conditions associated with abnormal vesicle  
CC trafficking, such as allergies, asthma, urticaria and autoimmune  
CC haemolytic anaemia. Anti-PRAM antibodies may be used as antagonists, as  
CC a targeting or delivery mechanism for bringing pharmaceutical agents,  
CC into cells or tissues expressing PRAM and for diagnosis of PRAM-related  
CC disorders. PRAM, its catalytic or immunogenic fragments are useful for  
CC drug screening using libraries of compounds. PRAM polynucleotides are  
CC useful for generating hybridisation probes useful in mapping the  
CC naturally occurring genomic sequences.  
XX  
SQ Sequence 142 AA;

Alignment Scores: 8.57e-59 Length: 142  
Pred. No.: 584.00 Matches: 104  
Score: 91.30% Conservative: 22  
Percent Similarity: 91.30%

Best Local Similarity: 75.36% Mismatches: 12  
Query Match: 39.22% Indels: 0  
DB: 21 Gaps: 0  
US-09-763-902b-10 (1-830) x AAV82323 (1-142)  
QY 116 TCCTGATTTTCAAGACCTATGTGATCAGGCGCTGCAGAGCTGCTGAGACTTTGTCAAT 175  
DB 4 SerLeuAspPheLysThrTyValAspGlnAlaLysArgAlaAlaGlnGluPheValAsn 23  
QY 176 GTCTACTACACACACCATGATTAAGCGCGCGCTGTGCTGCCCTGTACATGGGCACA 235  
DB 24 ILERYTYRGLTThrMetAspLysArgArgAlaGlnAlaLeuThrArgLeuTyLeuAspLys 43  
QY 236 GCCACCGCTGGCTGGATGCAATGCAATGCTTTCAGACACAGAACTTGTAGAGATTGTTT 295  
DB 44 AlathrLeuThrPheAsnGlnAspAlaValSerIleuAspAlaLeuAsnAspPhe 63  
QY 296 GAATGTTCCTTCCAGCGAGTTCACAAATCAGCGCTGTAGACTCCAGCCTGTTCATGAT 355  
DB 64 AspThrLeuProSerSerGlnPheGlnValAsnMetLeuAspCysGlnProValHisGlu 83  
QY 356 GAAGCCACACCAAGCCAGACCGCGCTCTGTGTATCATGTGTGATCAGTGAATTGAG 415  
DB 84 GlnAlaThrGlnSerGlnThrThrValLeuValAlaThrSerGlyThrValLysPheAsp 103  
QY 416 GGGAAACAACAAGCGAGCTTCAACCACTTCATCCGACCGCGCGCTTCACCCGAC 475  
DB 104 GlnAsnLysGlnHisPhePheAsnGlnAsnPheLeuThrAlaGlnSerThrProAsn 123  
QY 476 AACACAGTGTGGAAGATCGCAAGTACTGCTCCGCTTCCAGACTGGCGCAGC 529  
DB 124 AsnThrValTrpLysIleAlaSerAspCysPheArgPheGlnAspTrpSer 141  
RESULT 5  
ABB04459  
ID ABB04459 standard; Protein: 142 AA.  
XX  
AC ABB04459;  
XX  
DT 04-MAR-2002 (first entry)  
XX  
DE Human NTF2 associated protein 16.  
XX  
KW Human; NTF2 associated protein 16; cancer; HIV infection; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN CN133332-A.  
XX  
PD 19-SEP-2001.  
XX  
PF 15-MAR-2000; 2000CN-0114918.  
XX  
PR 15-MAR-2000; 2000CN-0114918.  
XX  
PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
XX  
PI Mao Y, Xie Y;  
XX  
DR WPI: 2002-042199/06.  
XX  
DR N-PSDB: ABA05754.  
XX  
PT Polypeptide-human NTF2 associated protein 16 and polynucleotide  
PT encoding it  
XX  
PS Claim 1; Page 26(Disclosure); 32pp; Chinese.  
XX  
CC The present invention provides the protein and coding sequences of human  
CC NTF2 associated protein 16. The sequences can be used in the treatment of  
CC cancer and HIV infection, as well as other diseases. The present sequence  
CC is the protein of the invention.



PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246509.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249269.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
DR WPI: 2001-488783/53.  
DR N-PSDB: AAS26214.  
XX  
XX New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives  
XX  
XX Claim 11; SEQ ID NO 1180; 980pp; English.  
XX  
XX The invention relates to isolated nucleic acid molecules and their

CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunoassays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. Rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention.

## Alignment Scores:

Pred. No.:	3,49e-58	Length:	162
Score:	579.00	Matches:	103
Percent Similarity:	91.18%	Conservative:	21
Best Local Similarity:	75.74%	Mismatches:	12
Query Match:	38.89%	Indels:	0
DB:	22	Gaps:	0

US-09-763-902b-10 (1-830) x AAUI6227 (1-162)

QY	122	GATTTCAGACCTATGTGATCAGAGCCCTGCAGAGCTGCTGAGATTGTCAATGCTAC	181
DB	26	AspPheLysTrpTyrValAspGlnAlaCysArgAlaGluGluPheValAsnIleTyr	45
QY	182	TACACACCATGATGATTAACGGCGGCTTGTGTCGCCCTGTCATGAGGACAGCCACC	241
DB	46	TyrGlnTrpMetAspLysArgArgArgAlaLeuThrArgLeuTyrLeuAspLysAlaThr	65
QY	242	CTGCTCTGGAATGGCATGCTGTTTCAGAGACAGATCCCTGAGTGTATTTGAATG	301
DB	66	LeuIleTrpAsnGlyAlaValSerGlyLeuAspAlaLeuAsnAspPheSer	85
QY	302	TTCCTTCAGAGCTTCCAAATCAGCTGTAGTCCAGCCCTGTCATGATGATGACCC	361
DB	86	LeuProSerSerGluPheGlnValAsnMetLeuAspCysGlnProValHisGlnIleAla	105
QY	362	ACACCAAGCAGACAGCAGCTGCTGTTTCATCTGTGATGATGATGATGATGAGGAGAC	421
DB	106	ThrGlnSerGlnThrThrValLeuValValThrSerGlyThrValLysPheAspLysn	125
QY	422	AAACAACGGAGCTTCAACAGCAACTTCACTGACCGCCAGAGCCCTCCAGCAACACA	481
DB	126	LysGlnHisPhePheAsnGlnAsnPheLeuThrAlaGlnSerThrProAsnAsnThr	145
QY	482	GTCGTGGAAGATGCAAGTGAAGTCTCCGCTTCCAGAGCTGAGCCAGC	529
DB	146	ValTrpLysIleAlaSerAspCysPheArgPheGlnAspTrpSerSer	161

## RESULT 7

ABB65017 standard; Protein: 133 AA.

ABB65017;

26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 21843.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW

KM pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN W0200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US092331.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PERK) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PMD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
DR N-PSDB; ABL09120.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions  
PS Disclosure; SEQ ID NO 21843; 21pp + Sequence Listing; English.  
XX  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB161737-AB162072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
XX  
SQ Sequence 133 AA:  
  
Alignment Scores:  
Pred. No.: 1,04e-26 Length: 133  
Score: 308.00 Matches: 55  
Percent Similarity: 61.65% Conservative: 27  
Best Local Similarity: 41.35% Mismatches: 47  
Query Match: 20.69% Indels: 4  
DB: 22 Gaps: 1  
  
US-09-763-902B-10 (1-830) x ABB65017 (1-133)  
  
QY 122 GATTTCAGACCTATGTGATCGAGCGCTCGAGAGCTGAGAGTTTGCATGTCTAC 181  
DB 4 AspleubylsalalysalvalguserCysAlaArgThrAlaAspThrArgLeuTyr 23  
QY 182 TACACACACCTGATTAAGCGCGCGCTTCCTCCCGCTGACATGGGACACAGCAC 241  
DB 24 TyralSerValAspSnaArgArgGlnGlnIleGlyArgLeuTyrLeuAspAsnAlaThr 43  
QY 242 CTGCTGTGAATGCAATGCTGTTTCAAGACAGAAATCCCTGAGTGTGATTTTGAATG 301  
DB 44 LeuPrtSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 63  
QY 302 TTGCTTTCCAGAGTTCACAAATCAGCGCTGAGTACGCGCTGCTGATGATGAAGC 361  
DB 64 LeuPrtSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 83  
QY 362 ACACACACGACGACGACGCTGCTGTTGATCTGATGATGATGATGATGATGATGATG 421  
DB 84 ValSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 103  
QY 422 AAACACAGGACCTTCAACGAACTTCATCTTCAAGCGCGCGCTGATGATGATGATGAT 481  
DB 104 GlnLeuArgGlySpheGlnGlnThrPheIleValThrAlaGlu-----AsnAsp 119

QY 482 GTGTCGAGATGCGCAAGTACTGCTTCGCGCTTCACAGAC 520  
DB 120 LysTrpLysValValSerAspCysTyrArgGlnGln 132  
  
RESULT 8  
ID AAG12073 standard; Protein; 125 AA.  
AC AAG12073;  
XX  
XX  
DT 17-OCT-2000 (first entry)  
XX  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 11042.  
XX  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX  
PN EP1033405-A2.  
XX  
XX  
PD 06-SEP-2000.  
XX  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
PR 25-FEB-1999; 99US-0121825.  
XX  
PR 05-MAR-1999; 99US-0123180.  
XX  
PR 09-MAR-1999; 99US-0123548.  
XX  
PR 23-MAR-1999; 99US-0125788.  
XX  
PR 25-MAR-1999; 99US-0126264.  
XX  
PR 29-MAR-1999; 99US-0126785.  
XX  
PR 01-APR-1999; 99US-0127462.  
XX  
PR 06-APR-1999; 99US-0128234.  
XX  
PR 08-APR-1999; 99US-0128714.  
XX  
PR 16-APR-1999; 99US-0129845.  
XX  
PR 19-APR-1999; 99US-0130077.  
XX  
PR 21-APR-1999; 99US-0130449.  
XX  
PR 23-APR-1999; 99US-0130510.  
XX  
PR 28-APR-1999; 99US-0130891.  
XX  
PR 30-APR-1999; 99US-0131449.  
XX  
PR 30-APR-1999; 99US-0132048.  
XX  
PR 04-MAY-1999; 99US-0132407.  
XX  
PR 05-MAY-1999; 99US-0132484.  
XX  
PR 06-MAY-1999; 99US-0132485.  
XX  
PR 06-MAY-1999; 99US-0132486.  
XX  
PR 06-MAY-1999; 99US-0132487.  
XX  
PR 07-MAY-1999; 99US-0132663.  
XX  
PR 11-MAY-1999; 99US-0134256.  
XX  
PR 14-MAY-1999; 99US-0134259.  
XX  
PR 14-MAY-1999; 99US-0134261.  
XX  
PR 14-MAY-1999; 99US-0134320.  
XX  
PR 18-MAY-1999; 99US-0134370.  
XX  
PR 19-MAY-1999; 99US-0134941.  
XX  
PR 20-MAY-1999; 99US-0135124.  
XX  
PR 21-MAY-1999; 99US-0135353.  
XX  
PR 24-MAY-1999; 99US-0135629.  
XX  
PR 25-MAY-1999; 99US-0136021.  
XX  
PR 27-MAY-1999; 99US-0136392.  
XX  
PR 28-MAY-1999; 99US-0136782.  
XX  
PR 01-JUN-1999; 99US-0137222.  
XX  
PR 03-JUN-1999; 99US-0137528.  
XX  
PR 04-JUN-1999; 99US-0137724.  
XX  
PR 07-JUN-1999; 99US-0138094.  
XX  
PR 08-JUN-1999; 99US-0138094.  
XX  
PR 10-JUN-1999; 99US-0138847.  
XX  
PR 14-JUN-1999; 99US-0139119.  
XX  
PR 16-JUN-1999; 99US-0139453.  
XX  
PR 16-JUN-1999; 99US-0139453.  
XX  
PR 17-JUN-1999; 99US-0139492.



DB 2 SerProLysGlyIleGluGluValAlaSerAla----- 13  
QY 146 GCCTCAGACAGCTGAGAGCTTTGTCAATGCTCTACACACCAGTGAATAGCGCGG 205  
DB 14 -----PheValAsnHisTyrIleHisLeuPheAspAsnSparG 26  
QY 206 CGTTGCTGTCCCGCTGTACATGGGACAGCCACCCGTGGTGAATGCGAATCCTGTT 265  
DB 27 SerSerLeuSerSerLeuTyrAsnProThrSerLeuLeuThrPheGluGlyGlnThrIle 46  
QY 266 TCAGACACAGAATCCCTTGAGTGAATTTTGAATGTCCTTCAGGAGTCCAA--- 322  
DB 47 TyrGlyValAlaAspAsnHisLeuSerAsnLysLeuLysGlnLeuProPheAspGlnHisHis 66  
QY 323 ---ATCACCGCTGTGACTGCGACGCTGTTTCATGATGAAGCACACACAGCAGCAGCAG 379  
DB 67 LeuIleSerThrValAlaAspSerGlnPro-----SerSerMetAlaGlyGlyCysGly 84  
QY 380 GTCTTGTGTCATCTGTGATCAGTGAAGTTGAGGG---ACCAACACAGGAGCTTC 436  
DB 85 IleLeuValAlaPheValSerGlySerIleGlnLeuHisGlyGluAspHisProLeuArgPhe 104  
QY 437 AACCAAGAACTTCATCTGACGCGCCAGCGCTTCACCCACACACAGTGTGAAGATGCA 496  
DB 105 SerGlnValTyrLeuLeu----- 110  
QY 497 AGTGACTGCTCCGCTTCAG 517  
DB 111 SerIleCysAsnArgPheGlu 117

RESULT 9  
AAG05145  
ID AAG05145 standard; Protein; 153 AA.  
XX AAG05145;  
AC AAG05145;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1441.  
XX  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0133485.  
PR 06-MAY-1999; 99US-0133486.

PR 06-MAY-1999; 99US-0133487.  
PR 07-MAY-1999; 99US-0133863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0138119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139494.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.

```

PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145991.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

```

```

PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 8,14e-05 Length: 153
Score: 120.00 Matches: 43
Percent Similarity: 46.31% Conservative: 26
Best Local Similarity: 28.86% Mismatches: 58
Query Match: 8,06% Indels: 22
DB: 21 Gaps: 7

US-09-763-902b-10 (1-830) x AAG05145 (1-153)
OY 86 AGCCCTCTCCATAGAACACAGATGCATCTGTGATTTCAAGACCTATGTGATCAG 145
DB 12 SerSerAlaSerLeuSerLeuLysMetSerLeuMetLsp-----ProAspAla 27
OY 146 GCCTGACAGCTGCTGAGAGATTGTGCATGTCTACATGCACACACATGATAGCGCG 205
DB 28 ValSerLysAla-----PheValGlnIstIstYrYrSerThrPheAspThrAsnArg 44
OY 206 CGTTGCTGTCCSCCTGTACATGAGGACACACACACCTGTGTGATGCAATGCTGTT 265
DB 45 ValGlyLeuAlaGlyLeuTyrGlnGlnIstAlaSerMetLeuThrPheGlnGlyGlnLysIle 64
OY 266 TCAGGACAGAAATCCTTGACAGTGTGTTTGAATGTGCT-----TCAGGAGAGTTC 319
DB 65 GlnGlyValGlnSerIleValAlaLysLeuThrSerLeuProPheGlnGlnLysLysIle 84
OY 320 CAATGACGCGTGTAGATGCGACGCTGTTCATGATGAGAACACACACACGCG 379
DB 85 HistLeuSerThrValAspCysGlnPro-----SerGlyProAlaSerLys 99
OY 380 GTCCCTGTGTGATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 436
DB 100 MetLeuValPheValSerGlyAsnLeuGlnLeuAlaGlyLysIstAlaLysLeuPhe 119
OY 437 AACCCAGAACTTCATCTG-----ACGCGCCAGGCTTACCCAGACACAGATGTGAGAG 490
DB 120 SerGlnMetPheHisLeuMetSerProThrProGlnIstSer-----PheTyr 134
OY 491 ATCCGAGTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 517
DB 135 ValPheAsnAspIlePheSerThrArg 143.

RESULT 10
ABB67404
ID ABB67404 standard; Protein: 688 AA.
AC ABB67404;
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster polypeptide SEQ ID NO 29004.
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
OS Drosophila melanogaster.
XX PN WO200171042-A2.

```

```
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX AC ABB66052;
XX XX ABB66052;
XX DT 26-MAR-2002 (first entry)
XX XX
XX DE Drosophila melanogaster polypeptide SEQ ID NO 24948.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL101507.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 29004; 21pp + Sequence Listing; English.
XX XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABBS7737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 688 AA;

Alignment Scores:
Pred. No.: 0.000225 Length: 688
Score: 119.00 Matches: 37
Percent Similarity: 45.90% Conservative: 19
Best Local Similarity: 30.33% Mismatches: 54
Query Match: 7.99% Indels: 12
DB: 22 Gaps: 5

US-09-763-902b-10 (1-830) x ABB67404 (1-688)
QY 164 GAGTTTGTCAATCTCTACTACACACCATGGATAAGCGGGGTTGTCTGCCCGCTG 223
Db 17 GlupheValArgGlnTyrThrLeuLeuAsnLysAlaProAsnHisLeuHisArgPhe 36
QY 224 TACATGGGCACACCCACCTGGTC---TGGATGGCAATGCTGTTTCAGGACAAGATCC 280
Db 37 TyrAsnHisAsnSerSeryrilleHisGlyGluSerLysLeuValValGlyGlnArgGlu 56
QY 281 TTGAGTGAGTCTTTTGAATGTTCCCTCCAGCGAG-----TTCCAATACCGTGGTA 334
Db 57 IleHisAsnArgIleGlnLeuAsnPheAsnAspCysHisAlaLysIleSerGlnVal 76
QY 335 GACTGGCCAGCTTTTCATGATGAAGCCACACCAAGCCAGCAGCTGCTGTGTGTCATC 394
Db 77 AspAlaGlnAlaThrLeuGlyAsnGly-----ValValValGlnVal 90
QY 395 TGTGATCATGTAAGTTTGGGGGAACAACACGCGGACTTCAACGAGACTTCATCTCTG 454
Db 91 ThrGlyGluLeuSerAsnAspGlyGlnProMetArgArgPheThrGlnThrPheValLeu 110
QY 455 ACCGCCACGCCCTCACCAGCACACAGTGTGGAAGATCGCAAGTACTGCTTCCGCTTC 514
Db 111 AlaAlaGln---SerProLysLys-----TyrTyrValHisAsnAspIlePheArgTyr 127
QY 515 CAGGAC 520
Db 128 GlnAsp 129

RESULT 11
ABB66052
ID ABB66052 standard; Protein; 690 AA.
XX AC ABB66052;
XX XX ABB66052;
XX DT 26-MAR-2002 (first entry)
XX XX
XX DE Drosophila melanogaster polypeptide SEQ ID NO 24948.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR N-PSDB; ABL10155.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Disclosure; SEQ ID NO 24948; 21pp + Sequence Listing; English.
XX XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABBS7737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 690 AA;

Alignment Scores:
Pred. No.: 0.000225 Length: 690
Score: 119.00 Matches: 37
Percent Similarity: 45.90% Conservative: 19
Best Local Similarity: 30.33% Mismatches: 54
Query Match: 7.99% Indels: 12
DB: 22 Gaps: 5

US-09-763-902b-10 (1-830) x ABB66052 (1-690)
QY 164 GAGTTTGTCAATCTCTACTACACCATGGATAAGCGGGGTTGTCTGCCCGCTG 223
Db 19 GlupheValArgGlnTyrThrLeuLeuAsnLysAlaProAsnHisLeuHisArgPhe 38
QY 224 TACATGGGCACACCCACCTGGTC---TGGATGGCAATGCTGTTTCAGGACAAGATCC 280
Db 39 TyrAsnHisAsnSerSeryrilleHisGlyGluSerLysLeuValValGlyGlnArgGlu 58
QY 281 TTGAGTGAGTCTTTTGAATGTTCCCTCCAGCGAG-----TTCCAATACCGTGGTA 334
Db 59 IleHisAsnArgIleGlnLeuAsnPheAsnAspCysHisAlaLysIleSerGlnVal 78
```



QY	335	GAGTCGACGCCCTTCATCATATGATTAAGCCACACCAAGCCACACCGAGCTCCCTGTGTGCATC	394
Db	79	AspAlaGlnAlaIatThLeuGlyasnly-----ValValValGlnVal	92
QY	395	TGTGCATGAGTGAAGATTGTGAGGGAACAACAACGGAGACTTCACCAAGACTTCATCCTG	454
Db	93	ThrglyGulLeuSerAsnAspIylGlnPromeTarGArgPheThrGlnThrPheValLeu	112
QY	455	ACCGCCCAAGGCGCTCACCCACACACAGTGTGGAGAGATGCGAAGTACTGCTCCGCTTC	514
Db	113	AlaAlaGln--SerProIysIys-----TyrTyrValHisAsnAspIlePheArgTyr	129
QY	515	CAGGAC	520
Db	130	Glnasp	131
RESULT 12			
ID	AA08682	standard: Protein; 123 AA.	
XX	AA08682;		
DT	17-OCT-2000	(first entry)	
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 6317.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
XX	termination sequence.		
OS	Arabidopsis thaliana.		
PN	EP1033405-A2.		
PD	06-SEP-2000.		
XX			
XX	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	04-MAY-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134376.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	

```

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154038.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Alignment Scores:
Pred. No.: 0.000143
Score: 117.50
Percent Similarity: 47.97%
Best Local Similarity: 30.08%
Query Match: 7.89%
DB: 21
Gaps: 5

US-09-763-902b-10 (1-830) x AAG08682 (1-123)

QY 158 GCTGAGGACTTTGTCTACTACTACACCATGATAGCGGCGTTGCTGCC 217
      ..... ||||| ||||| |||||
DB 7 SerIysAlaPheValGluHisItyrItyrSerIthPheAspThrAsnArgValGlyLeuAla 26

QY 218 CGCGTGTACATGGCCAGCCACCCTGCTGGAATGCAATGCTTTCAGCAAGA 277
      ||||| ||||| ||||| |||||
DB 27 GlyLeuTyrGlnGluHisSerMetLeuThrPheGluGlyGlnLysIleGlnIyValGln 46

QY 278 TCCTTGAGTGAAGTTTGAATGTTGCTT-----TCCAGCGAGTTCCAAATCAGCTG 331
      ||||| ||||| ||||| |||||
DB 47 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnCysIshHisIleSerThr 66

QY 332 GTAGACTGCCAGCTTTCATGATGAGCACACCAACCAACGCGCTTGTTC 391
      ||||| ||||| ||||| |||||
DB 67 ValAspCysGlnPro-----SerGlyProAlaSerIyMetLeuValPhe 81

QY 392 ATCTGTGATCAGTGAAGTTTGAAGGAGCAACCAACGCG---GACTTCAACGAGACTTC 448
      ..... ||||| ||||| |||||
DB 82 ValSerGlyAsnLeuGlnHisIleValGlnGlnHisIleValGlnLysPheSerGlnMetPhe 101

QY 449 ATCTGTG-----ACCGCCAGCGCTTCACCCAGCAACAGTGTGAGATCGCAAGTGC 502
      ||||| ||||| ||||| |||||
DB 102 HisLeuMetProThrProGlnGlnIySer-----PheTyrValPheAsnAsp 116

QY 503 TGCTTCCGC 511
      |||||
DB 117 IlePheArg 119

RESULT 13
AAG08681
ID AAG08681 standard; Protein; 126 AA.
XX AC
XX AAG08681;
XX
DB 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6316.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.

```

PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132853.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147933.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158363.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 0.000144 Length: 126  
Score: 117.50 Matches: 37  
Percent Similarity: 47.97% Conservative: 22  
Best Local Similarity: 30.08% Mismatches: 49  
Query Match: 7.89% Indels: 15  
DB: 21 Gaps: 5

US-09-763-902b-10 (1-830) x AAG08681 (1-126)

QY 158 GCGAGAGATGTGTCATGTCATGACACCATGATGATAGCGGCGCTTGTCTCC 217  
DB 10 SerLeuAlaPheValAluHisTyrTyrSerThrPheAsnArgValGlyLeuAla 29  
QY 218 GCGCTTACATGGGCGACACGACCTGTCTGATGCAATGCTTTCCAGCAAGA 277  
DB 30 GlyLeuTyrGlnIleuLaserMetLeuThrPheGlnGlyIleGlnIleGlnIle 49  
QY 278 TCTTTCAGTGAGTTTGTGAAATGTTGCTT-----TCCAGCGATCCAAATCAGCG 331  
DB 50 SerLeuValAlaLysLeuThrSerLeuProPheGlnGlnIleCysLysHisIleSerThr 69  
QY 332 GATGCTGCCAGCCTGTTCATGATGACGACCAAGCAGCAGCGGCTTGTGTC 391  
DB 70 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 84  
QY 392 ATCTGTGATCACTGAAGTTTGAAGGGAACAACAACG--GACTTCAACCAAGACTTC 448  
DB 85 ValSerGlyAsnLeuGlnIleuAlaGlyGlnIleuHisAlaLeuLysPheSerGlnMetPhe 104  
QY 449 ATCTGTG-----ACGCGCCAGCGCTTCACCGACAGCAAGTGTGAGAGATGCAAGTGC 502  
DB 105 HisLeuMetProThrProGlnGlnIleSer-----PheTyrValPheAsnAsp 119  
QY 503 TGCTTCCGC 511  
DB 120 IlePheArg 122

## RESULT 14

AAG05147 standard; Protein: 131 AA.

AAG05147:

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 1443.

KX Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.

XX Arabidopsis thaliana.  
OS EPI033405-A2.  
XX 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139459.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0148171.  
PR 10-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152263.  
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 0.000147  
Score: 117.50  
Percent Similarity: 48.008  
Best Local Similarity: 29.608  
Query Match: 7.898  
DB: 21  
Gaps: 5

US-09-763-902B-10 (1-830) x AAG05147 (1-131)

QY 158 GCGTGGAGAGTTGTCATGTCCTACACCATGATTAAGGGGGGCTTTCGTGTC 217  
DB 7 SerLyAlaPheValIuhIstYrSerThnPheSphrsmkYvalGlyLeuAla 26  
QY 218 CGCCTGTACATGGGACACACCCCTGCTGGAATGGCAATCCTTTTCAGACAGAA 277  
DB 27 GlyLeuTyrgIngluAlaSerMetLeuThrPheGluGlyInlYsIleGInglYalGln 46  
QY 278 TCCTTGAGTACGTTTTCGAATGTCCTT-----TCCAGCAGTTCGAATCAGCGTG 331  
DB 47 SerIleValAlaLysLeuThrSerLeuProPheGInGInCysLysIshIstIleSerThn 66  
QY 332 GTAGACTGCACGCTTCATGATGAAGCAGCAGCAGCAGCGGTCCTGTGTC 391  
DB 67 ValAspCysGlnPro-----SerGlyProAlaSerGlyLeuValPro 81  
QY 392 ATCTGTGAGATCAGTGAAGTTTGAAGGAGAACAAACAGCG---GACTTCACAGCACTTC 448  
DB 82 ValSerClyAsnLeuGlnLeuAlaGlyGluYshIshIleValLeuLysPheSerGlnMetPhe 101

OY 449 ATCTG-----ACGCGCCAGCGCTCACCACACACAGTGTGAGATCCAGTGAC 502  
Db 102 HslsleuMetProthProclnclnclser-----PhetYrYalpheasnap 116  
OY 503 TCGTTCGCTCCAG 517  
Db 117 IlePheSerTrpArg 121  
RESULT 15  
AAG05146  
ID AAG05146 standard: Protein: 134 AA.  
AC AAG05146;  
XX  
XX 17-OCT-2000 (first entry)  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1442.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX Arabidopsis thaliana.  
OS  
XX  
XX EP103405-A2.  
PN  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148365.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 13-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:

Pred. No.: 0.000149  
 Score: 117.50  
 Percent Similarity: 48.008  
 Best Local Similarity: 29.608  
 Query Match: 7.898  
 DB: 21

Length: 134  
 Matches: 37  
 Conservative: 23  
 Mismatches: 50  
 Indels: 15  
 Gaps: 5

US-09-763-902b-10 (1-830) x AAC05146 (1-134)  
 QY 158 GCTGAGAGTTTGTCAATGCTACTACACCATGATAGCGCGCTTGTCTCC 217  
 Db 10 SerIysAlaPheValIGluHISrYrYrSerThrPheAspHrsmAtgValIGlyLeuAla 29  
 QY 218 CGCCTGTACATGGGACACACCGCCACCTGCTGTGANTGCATGCTGTTTCAG\$ACAGAA 277  
 Db 30 GlyLeuYrclngIngluAlaSerMetLeuThrPheclnuglIngluYsIleGlnGlyValGln 49  
 QY 278 TCCTTGAGTGAAGTTTGTGAAGTTCCTT-----TCACGCGATTCCAAATCAGCGTGC 331  
 Db 50 SerIleValAlaYsIleuThrSerLeuProPheGlnGlnYsIshISrThr 69  
 QY 332 GTAGACTGCCAGCCTGTTCATGATGAAGCCACACACACAGCAGCAGCGCTTGTCTC 391  
 Db 70 ValAspCysGlnPro-----SerIlyProAlaSerGlyMetLeuValPhe 84  
 QY 392 ATCTGTGATCAGTGAAGTTTGAAGGGAACAAACACAGG---GACTTCACACAGACTTC 448  
 Db 85 ValSerGlyAsnLeuGlnLeuAlaGlyGluYsHISrAlaLeuYsPheSerGlnMetPhe 104  
 QY 449 ATCCTG-----ACCGCCACAGCCTTCACCCAGCAACACAGTGTGAAGATCGCAAGTGAC 502  
 Db 105 HisLeuMetProThrProGlnGlySer-----PheYrValPheAsnAsp 119  
 QY 503 TGCCTCCGCTTCCAG 517  
 Db 120 IlePheSerTrpArg 124

Search completed: June 24, 2003, 19:31:03  
 Job time : 43.6938 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 19:27:58 ; Search time 10.711 Seconds  
(without alignments)  
4559.976 Million cell updates/sec

Title: US-09-763-902B-10  
Perfect score: 1489  
Sequence: 1 tgaacttgccattcactg.....ataatataataacacatg 830

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues  
Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_n2p.model -DEV=xlp  
-O/cgn2.1/USPTO.spool/US09763902/rnat\_24062003.130342.12495/app\_query.fasta.1.2254  
-DB=Issued Patents.AA -OFMT=fasta -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdf  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09763902.cgn1.1.42.0/rnat\_24062003.130342.12495 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-FEAPEXT=7 -Ygapop=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database: Issued Patents.AA:  
1: /cgn2.6/prodata/1/aa/5A.COMB.pep.\*  
2: /cgn2.6/prodata/1/aa/5B.COMB.pep.\*  
3: /cgn2.6/prodata/1/aa/5A.COMB.pep.\*  
4: /cgn2.6/prodata/1/aa/5B.COMB.pep.\*  
5: /cgn2.6/prodata/1/aa/PCTUS.COMB.pep.\*  
6: /cgn2.6/prodata/1/aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	90	6.0	466	US-08-836-791-9	Sequence 9, Appl1
2	88.5	5.9	671	US-08-858-207A-301	Sequence 101, App
3	81.5	5.5	498	US-08-660-963-12	Sequence 32, Appl
4	80	5.4	1276	US-08-222-616-24	Sequence 24, Appl
5	80	5.4	1276	US-08-446-648-24	Sequence 24, Appl
6	80	5.4	1276	PCT-US95-04228-24	Sequence 24, Appl
7	78.5	5.3	492	US-09-413-814-72	Sequence 72, Appl
8	78.5	5.3	2972	US-09-579-181-2	Sequence 2, Appl1
9	78.5	5.3	3118	US-09-579-181-1	Sequence 1, Appl1
10	78	5.2	480	US-08-570-227A-2	Sequence 2, Appl1
11	78	5.2	480	US-09-077-091-2	Sequence 2, Appl1
12	78	5.2	1098	US-08-290-301-82	Sequence 82, Appl

13	78	5.2	1098	US-09-013-598-82	Sequence 82, Appl
14	77	5.2	838	US-08-216-260-4	Sequence 4, Appl1
15	76.5	5.1	766	US-08-846-762-5	Sequence 5, Appl1
16	76	5.1	301	US-09-231-227-4	Sequence 4, Appl1
17	75	5.0	521	US-08-682-847-4	Sequence 4, Appl1
18	75	5.0	752	US-09-402-929-6	Sequence 6, Appl1
19	74.5	5.0	217	US-09-602-543-5	Sequence 5, Appl1
20	74.5	5.0	235	US-09-602-543-4	Sequence 4, Appl1
21	74.5	5.0	253	US-09-154-750A-83	Sequence 83, Appl
22	74.5	5.0	1081	US-09-369-364A-17	Sequence 17, Appl
23	74.5	5.0	1251	PCT-US95-02251-3	Sequence 3, Appl1
24	74.5	5.0	1252	US-08-199-780-3	Sequence 3, Appl1
25	74.5	5.0	1252	US-08-316-650-3	Sequence 3, Appl1
26	73.5	4.9	434	US-09-540-245A-19	Sequence 19, Appl
27	73.5	4.9	1082	US-09-336-643A-20	Sequence 20, Appl
28	73.5	4.9	3066	US-08-952-127-12	Sequence 12, Appl
29	73	4.9	310	US-08-943-600A-3	Sequence 3, Appl1
30	73	4.9	502	US-08-278-635B-7	Sequence 7, Appl1
31	73	4.9	502	US-08-464-258B-7	Sequence 7, Appl1
32	73	4.9	502	US-08-471-961-7	Sequence 7, Appl1
33	73	4.9	688	US-08-941-445A-11	Sequence 11, Appl
34	72.5	4.9	587	US-08-899-578-2	Sequence 2, Appl1
35	72.5	4.9	801	US-08-725-012-2	Sequence 2, Appl1
36	72.5	4.9	899	US-09-422-936-71	Sequence 71, Appl
37	72.5	4.9	902	US-09-644-600-10	Sequence 10, Appl
38	72.5	4.9	1037	US-09-428-711A-21	Sequence 21, Appl
39	72.5	4.9	1083	US-09-600-776-2	Sequence 2, Appl1
40	72.5	4.9	1083	US-09-343-494-1	Sequence 1, Appl1
41	72	4.8	163	US-08-727-688-25	Sequence 25, Appl
42	72	4.8	297	US-08-651-136C-4	Sequence 4, Appl1
43	72	4.8	297	US-09-229-811A-4	Sequence 4, Appl1
44	72	4.8	519	US-09-172-841-55	Sequence 55, Appl
45	72	4.8	1754	US-07-745-206A-13	Sequence 13, Appl

## ALIGNMENTS

RESULT 1  
US-08-836-791-9  
Sequence 9, Application US/08836791  
Patent No. 5886150  
GENERAL INFORMATION:  
APPLICANT: Duchesne, Marc  
APPLICANT: Faucher, Didier  
APPLICANT: Parker, Fabienne  
APPLICANT: Schweighofer, Fabien  
APPLICANT: Tocque, Bruno  
TITLE OF INVENTION: Peptides Capable of Binding to the GAP  
TITLE OF INVENTION: Protein SH3 Domain, Nucleotide Sequences Coding Thereof  
TITLE OF INVENTION: and Preparation and Use Thereof  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Road, Mailstop 3043  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/836,791  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94/13955  
FILING DATE: 22-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 95/05753  
FILING DATE: 16-MAY-1995



## PRIOR APPLICATION DATA:

APPLICATION NUMBER: NO FR95/01539  
FILING DATE: 22-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Fehlner Esq., Paul F.  
REGISTRATION NUMBER: 35,135  
REFERENCE/DOCKET NUMBER: ST94083G1-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 454-3839  
TELEFAX: (610) 454-3808  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 466 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-791-9

## Alignment Scores:

Pred. No.:	0.0524	Length:	466
Score:	90.00	Matches:	35
Percent Similarity:	44.09%	Conservative:	21
Best Local Similarity:	27.56%	Mismatches:	57
Query Match:	6.04%	Indels:	14
DB:	2	Gaps:	5

US-09-763-902B-10 (1-830) x US-08-836-791-9 (1-466)

QY 164 GAGTGTGCAATGCTGCTACACACACACATGATACGGGGCGGCTTGTCTCCCGCTG 223  
DB 14 GAluPhaValaGlnAlaGlnTyrTrpThrLeuLysGlnAlaProAspMetLeuHisArgPhe 33  
QY 224 TACATGGGACAGACACCTGCTGTGTAATGCC-----AAT 259  
DB 34 TyGlyLysAsnSerSerTyrValHisGlyGlyLeuAspSerAsnLysProAlaAsp 53  
QY 260 GCGGTTCAGGACAAATCCTTGAGTGAAGTTTGAATGTTGCTTCACAGGATTC 319  
DB 54 AlaValTyrGlyGlnLys-----GluLeuHisArgLysValMetSerGlnAsnPhe 70  
QY 320 CAATAGCGGTGTGTAATGCGCGCTGTCATGATGAGGCGCACCAACAGCCAGACAG 379  
DB 71 ThrAsnCyHisThrLysIleArgHisValAspAlaHisAlaThrLeuAsnAsp---Gly 89  
QY 380 GTCTGTGTGTCATCTGTGATCAGTGAAGTTTGAGGGGACAAACAGGACTTCAC 439  
DB 90 ValValValGlnValMetLysLeuLysSerAsnAsnGlnAlaLeuArgArgPheMet 109  
QY 440 CAGAACTTCATCTGACCGCCAGGCGCTCACCAGCAACAGATGTGGAATCCGCAAGT 499  
DB 110 GlnThrPheValLeuAlaProGluGlySerValAlaAsnLysPheTyr---ValHisAsn 128  
QY 500 GACTGCTTCGCTTCAGGAG 520  
DB 129 Asp---IleArgTyrGlnAsp 134

## RESULT 2

US-08-858-207A-301  
Sequence 301, Application US/08858207A  
Patent No. 6348328  
GENERAL INFORMATION:

APPLICANT: Black, Michael  
APPLICANT: Hodgson, John  
APPLICANT: Knowles, David  
APPLICANT: Nicholas, Richard  
APPLICANT: Stodola, Robert  
TITLE OF INVENTION: No. 6348328el Compounds  
NUMBER OF SEQUENCES: 552  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia

STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,207A  
FILING DATE: 09-MAY-1997  
CLASSIFICATION: 435

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/017670  
FILING DATE: 14-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmil, Edward R.  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50475  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:

INFORMATION FOR SEQ ID NO: 301:

SEQUENCE CHARACTERISTICS:

LENGTH: 671 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6348328e

US-08-858-207A-301

## Alignment Scores:

Pred. No.:	0.0962	Length:	671
Score:	88.50	Matches:	59
Percent Similarity:	36.43%	Conservative:	43
Best Local Similarity:	21.07%	Mismatches:	85
Query Match:	5.94%	Indels:	93
DB:	4	Gaps:	14

US-09-763-902B-10 (1-830) x US-08-858-207A-301 (1-671)

QY 8 TGGCATCAGCGTGGCTCTCTTCAACCTTCCCTGACGCCCTGTTCCCAAGCA 67  
DB 389 TrpHisAspTyrAsnProAsnGlnValLeuIleValThrProGlnTyrLeuGluArg 408  
QY 68 GAGGAATATCCCTGGTGGAGCCCTCTCCATGAAACAGAGATGGCATGTGGAT--- 124  
DB 409 GlnAsnIlePro-----ValAspThrThrIleGluGlnLysMetAsnHisLeuAspVal 426  
QY 125 -----TTCAAGCATATGTGATCAGGCGCTGACAGAGCTGCTAGAG---TTTGCAAT 175  
DB 427 GlyLysPheValLeuLeuLeuProGluHisLeuArgSerGluGlnHisTyrLysSer 446  
QY 176 GCTGAC-----TACACCATGATTAAGGGCGGCGCTTG--- 211  
DB 447 ValPheGluAspAspLeuThrSerArgMetSerSerGlnAspIleArgGlnMetThr 466  
QY 212 -----CTGTCCCGCTGTACATGAGGACAGCCAGCCCTGTGTAATGGCAATGCTGT 265  
DB 467 AlaThrValGlyTyrLeuGlnSerGlyGlnAspArgPheValTyrAsnThrProIle 486  
QY 266 TCAGGACAGAATCCTTGAAGTAGATTTTGAATGTTCCCTTCAGCAGATCCCAATC 325  
DB 487 SerTyrGlnGlnPheLeuLysAsp----- 494  
QY 326 ACCGTGTAGACTGCGCCCTGTTCATGATGAAGCAGCAGCAGCAGC----- 376  
DB 495 -----ProIleIleValIleThrProGlnSerThrGlyProGln 508  
QY 377 AGCGTCCTGTGTATCTGCTGATCAGTGAAGTTTGAGGGGACAAACAGCGGACTTC 436  
DB 509 SerIleLeuPheThrIle-----AspAla 516



## COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patin (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/222,616  
 FILING DATE: 4-APR-1994  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/00586  
 FILING DATE: 22-JAN-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 07/826935  
 FILING DATE: 22-JAN-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lee, Wendy M.  
 REGISTRATION NUMBER:  
 REFERENCE/DOCKET NUMBER: 821P2  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/952-1994  
 TELEFAX: 415/952-3881  
 TELEX: 910/371-7168  
 INFORMATION FOR SEQ ID NO: 24:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1276 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-222-616-24

## Alignment Scores:

Pred. No.: 143 Length: 1276  
 Score: 80.00 Matches: 59  
 Percent Similarity: 32.75% Conservative: 34  
 Best Local Similarity: 20.77% Mismatches: 86  
 Query Match: 5.37% Indels: 105  
 DB: 1 Gaps: 14

US-09-763-902b-10 (1-830) x US-08-222-616-24 (1-1276)

QY 33 CCTTACTTCCTGCGAGCCCTGTTCCCGCAGGAGCAAGAAATACCT---GGTGAGCC 89  
 DB 985 ProGlnYrProAlaGlyThrProHisProAlaGlyThrAlaSerProPheSerGlyAla 1004  
 QY 90 CTCCTTCATAGAACCAAGATGCGATCTGTGGATTTCAGAA-----CCTATGTGG 140  
 DB 1005 -----GluTrpGlyLeuThrGluAlaProSerProValProArgTrp 1018  
 QY 141 ATCAGGCTGCGAGAGCTGTCAGAGAGTTGTCAATGTCTACTACCCACCATGTAGC 200  
 DB 1019 IleAlaLeuAlaGlyGlyValAlaTrSer-----TrpGlnPhe 1030  
 QY 201 GCGGCGCTTCTCTGCTCCCGCTGTACATGGGACAGCCAGCCGCTGTGATGGCAATG 260  
 DB 1031 Gly-----GluThrGlyPheGlyGlySerAlaIle 1040  
 QY 261 CTGTTTACGACAGAAATCCTTGTAGTAGTATTTTGAATGTTCCCTCCAGCAGATTCC 320  
 DB 1041 IleGlyGlyGluAlaHisPro-----ProAlaThrSer 1051  
 QY 321 AAATACCGGTGTAGTATGCGAGCCTTATGATGTAAGCCACCAAGCCAGACCA--- 377  
 DB 1052 GlyAlaHiser-----ArgProAlaGlyValArgAlaProPhe 1062  
 QY 378 -----CGGTCTGTGTGTCATCTGTGATGTCAGTGAAGTTTGAAG 416  
 DB 1063 ProGlnAspTrpValProGluGluValGlyValAlaProSerHisSerGln----- 1078  
 QY 417 GGAACAACAAGCGGAGCTTACACCAAGATTCTGACCGCCAGG-----CCTCAC 470  
 DB 1079 -----ProProGlnValProProSer 1085  
 QY 471 CCAGCAACACAGTGTGGAGA---TCGCAAGTGTACTGCTTCCCTTCAGAGACTGGCCA 527

DB 1086 ProTrpValArgSerArgAlaGlyProLysArgValLeuProCysGlnLeuGlnSerGlyGly 1105  
 QY 528 GCTAGTGGCGGTGGCGAGAGGTCTCTTGTTCATTCAGCCCTAGCTCTGTAGCAAAATGC 587  
 DB 1106 AlaValProGlyGlyLys-----LysGlyCys 1114  
 QY 588 AAACCTGCACTCTCAAGGATGTGAGACACAGATTCTTCTGTGTGTCGGAGACACT 647  
 DB 1115 GlnGlyProValThrLysSerLeuGlyPheValAlaProThrCysCysHisAlaGln 1134  
 QY 648 GCAGACTCCAGCTGTCCGAGGTGAACCTC----- 677  
 DB 1135 ThrGlnSerPhePheSerLeuValAsnAlaProProAlaAlaAlaPheIleLeuLys 1154  
 QY 678 -----TTTGTGCTCAAGTTCTAGAGACTCCCTTCTGAAATATATCTGTT-- 726  
 DB 1155 ValPheGluPhe-CysPheTrpSerPhePheSerProPheProPheCysPhePheValle 1174  
 QY 727 -----TGTCAATAGTTTCTTTCATTAAGTAAGTAAGTTTCTATTCTTCT 770  
 DB 1174 upPhePheTrpArgProCysHisAsnPheValLeuGlnGlyThrCysPhePheMetalase 1194  
 QY 771 ACTGCCCGAG 780  
 DB 1194 rPheAlaGln 1197

RESULT 5  
 US-08-446-648-24  
 Sequence 24, Application US/08446648  
 Patent No. 6311302  
 GENERAL INFORMATION:  
 APPLICANT: Genentech, Inc.  
 APPLICANT: Bennett, Brian D.  
 APPLICANT: Goeddel, David.  
 APPLICANT: Lee, James M.  
 APPLICANT: Matthews, William  
 APPLICANT: Tsai, Siao Ping  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES  
 NUMBER OF SEQUENCES: 45  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Genentech, Inc.  
 STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Winpatin (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/446,648  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/222616  
 FILING DATE: 04-APR-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Lee, Wendy M.  
 REGISTRATION NUMBER: 40,378  
 REFERENCE/DOCKET NUMBER: P0821P3PCT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/225-1994  
 TELEFAX: 415/952-9881  
 TELEX: 910/371-7168  
 INFORMATION FOR SEQ ID NO: 24:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1276 amino acids  
 TYPE: Amino Acid  
 TOPOLOGY: Linear

US-08-446-648-24

## Alignment Scores:

Pred. No.:	1.43	Length:	1276
Score:	80.00	Matches:	59
Percent Similarity:	32.75%	Conservative:	34
Best Local Similarity:	20.77%	Mismatches:	86
Query Match:	5.37%	Indels:	105
		Gaps:	14

US-09-763-902B-10 (1-830) x US-08-446-648-24 (1-1276)

```

QY 33 CCTTACTCCCTGACAGCCCTGCTGCCAGGACAGAGAAATACCT---GGTGAGCC 89
DB 985 PROGLNTPRVALHGLYTHPRHLSPRHLSASPHALASERPROPHESERGLYALA 1004
QY 90 CTCCTTCATAGAACAGAGATGGCATCTGTGATTTCAAGA-----CCTATGTG 140
DB 1005 -----GluTPRGLYLeuThrgLUALPRoserProValPRoArgTRP 1018
QY 141 ATCAGCCCTGCAGAGCTGCTGAGAGATTGTCTACTACACACCATGATAGC 200
DB 1019 ILeAlaLeuAlaArgGLYAlaArgSer-----TTPGLNphe 1030
QY 201 GGGGGGGTTGCTGCCCGCTGACATGGGACAGCCAGCTGTGTGAATGGCAATG 260
DB 1031 GLY-----GluThrgLYPheGLYGLYSerAlaIle 1040
QY 261 CTGTTTCAGACAGAAATCCTTGATGAGTGTGTTTGAATGTGCTTCCAGCAGTTC 320
DB 1041 ILeGLYGLYGLYLeuSnhISPro-----ProAlaThSser 1051
QY 321 AATTCAGCCGTGAGATGCGACAGCTTTCATGATGAGCCACACACAGCAGCACA 377
DB 1052 GLYasnSer-----ArgProAlaValAlaArgAlaProPhe 1062
QY 378 -----CGGTCTGTGTGTCATCTGTGATGATGATGATGATGATGATGATG 416
DB 1063 PROGLNAsPTPRVALPRoGLNGLYGLYAlaPRoAlaPRoAlaPRoAlaPRo 1078
QY 417 GGAACAACAACGGGACTTCACACGAATTCATCTGACGCCAGG-----CTTCAC 470
DB 1079 -----ProProGLNValPRoProSer 1085
QY 471 CCAGCAACACAGTGTGAAGA---TCCCAAGTACTGCTTCCAGTACTGGGCCA 527
DB 1086 PRoTrPRVALHrSerArGArgPRoLYsArGLYValLeuPRoCysGLNLeuGLNserGLY 1105
QY 528 GCTGTGGGGGTGCAGAGCTCTTGTCTTCAATCAGCCCTAGCTGTGAGAGAAATGC 587
DB 1106 AlAValPRoGLYGLYs-----LYsGLYcys 1114
QY 588 AAACCTGACCTCTCAGAGATGTAGAACACAGATTCATTTCTGTGTGGGAGACACT 647
DB 1115 GLNGLYPRoVALHrLYsSerLeuGLYPhEVALPRoHrCysCysHIShISGLN 1134
QY 648 GCAGCTCCTGCTGCGAGGTGAACCT----- 677
DB 1135 ThrGLNserPhePheSerLeuValASnAlaPRoPRoAlaAlaAlaPheLeuLYs 1154
QY 678 -----TTTTGTGCTCAAGTCTTCAAGATCCCTTCTGTAATATACTGT-- 726
DB 1155 ValPheGLNphe-CysPheTrPserPhePheSerProPheProPheCysPhePheValLe 1174
QY 727 -----TGTCAATAGTCTCTTTCAAAGTAGTAACCTTTTCAATTTTCT 770
DB 1174 uphEphETyArGrPRoCysHISAsnPhEVALLeuGLNGLYHrCysPheThMeAlaSe 1194
QY 771 ACTTGCCAG 780
DB 1194 rPheAlaGLN 1197

```

RESULT 6

PCT-US95-04228-24

Sequence 24, Application PC/TUS9504228

## GENERAL INFORMATION:

APPLICANT: Genentech, Inc.  
 APPLICANT: Bennett, Brian D.  
 APPLICANT: Goeddel, David  
 APPLICANT: Lee, James M.  
 APPLICANT: Matthews, William  
 APPLICANT: Tsai, Siao Ping  
 APPLICANT: Wood, William I.  
 TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES  
 NUMBER OF SEQUENCES: 45  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Genentech, Inc.  
 STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080

## COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patin (Genentech)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/04228  
 FILING DATE:  
 CLASSIFICATION:

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/222616  
 FILING DATE: 04-APR-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Wendy M. Lee  
 REGISTRATION NUMBER: 00,000  
 REFERENCE/DOCKET NUMBER: 821P3PCT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/225-1994  
 TELEFAX: 415/952-9881  
 TELEX: 910/371-7168

## INFORMATION FOR SEQ ID NO: 24:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 1276 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 PCT-US95-04228-24

## Alignment Scores:

Pred. No.:	1.43	Length:	1276
Score:	80.00	Matches:	59
Percent Similarity:	32.75%	Conservative:	34
Best Local Similarity:	20.77%	Mismatches:	86
Query Match:	5.37%	Indels:	105
		Gaps:	14

US-09-763-902B-10 (1-830) x PCT-US95-04228-24 (1-1276)

```

QY 33 CCTTACTCCCTGACAGCCCTGCTGCCAGGACAGAGAAATACCT---GGTGAGCC 89
DB 985 PROGLNTPRVALHGLYTHPRHLSPRHLSASPHALASERPROPHESERGLYALA 1004
QY 90 CTCCTTCATAGAACAGAGATGGCATCTGTGATTTCAAGA-----CCTATGTG 140
DB 1005 -----GluTPRGLYLeuThrgLUALPRoserProValPRoArgTRP 1018
QY 141 ATCAGCCCTGCAGAGCTGCTGAGAGATTGTCTACTACACACCATGATAGC 200
DB 1019 ILeAlaLeuAlaArgGLYAlaArgSer-----TTPGLNphe 1030
QY 201 GGGGGGGTTGCTGCCCGCTGACATGGGACAGCCAGCTGTGTGAATGGCAATG 260
DB 1031 GLY-----GluThrgLYPheGLYGLYSerAlaIle 1040
QY 261 CTGTTTCAGACAGAAATCCTTGATGAGTGTGTTTGAATGTGCTTCCAGCAGTTC 320

```

```

Db      1041 IieglyglylualshisPro-----Proalathiser 1051
Oy      3321 AATAGCGCTGTATACACGCCAGCTGTTCATGATGAAAGCCAGACCAAGCCAGACCA 377
Db      1052 Glysenser-----ArgProArValArAlarprohe 1062
Oy      378 -----CGGTCCTGTGTGTATCTGTGAGTACAGTGAAGTTTGAG 416
Db      1063 ProGlnAPTPValPProGlnGluLysGluValProAsnIleSerGln----- 1078
Oy      417 GGACAAACAAAGGAGCTTCAACCAAGACTTCATCTGAGCGCCAGG-----CCTCAC 470
Db      1079 -----ProGlnValProProser 1085
Oy      471 CCAGACACAGAGTGTGAAGA---TCGCAGTACTGCTCCGCTCCAGAGACTGGGCGCA 527
Db      1086 ProTPValArGserArGArProLysArGValLeuProGysGlnLeuGlnSerGly 1105
Oy      528 GCTAGTGGGGTGGCAGAGGCTCTTCTTCTTCATTCAGCCCTAGCTGTAGAGAAATGC 587
Db      1106 AlaValProGlyGlyLys-----LysGlyCys 1114
Oy      588 AAACCTGCATCTCAAGAGTGTGAGAAACAGATTCATTTCTGTGTGGGAGACACT 647
Db      1115 GlnGlyProValThrLysSerLeuGlyPheValValProThrCysCysShisGln 1134
Oy      648 GCACATCCACTGTGCGGAGCTTGAACCT----- 677
Db      1135 ThrGlnSerPhePheSerLeuValAsnAlaProProAlaAlaPheIleLeuLys 1154
Oy      678 -----TTTTGTGCTCAAGTCTAGAGTCCCTTCTCAATATATACTGTT- 726
Db      1155 ValPheGlnPhe-CysPheTrpSerPhePheSerProPheProheCysPhePheVal 1174
Oy      727 -----TGTCATGTTCTTCTTCAAGTGAAGTAACTTTCTATTCTTCT 770
Db      1174 upPhePheTrpArGProCysShisAsnPheValLeuGlyLysThrcysPheThrMetaLase 1194
Oy      771 ACTTGCCAG 780
Db      1194 rPheAlaGln 1197

```

```

RESULT 7
US-09-413-814-72
Sequence 72, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloecker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hofle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 72
LENGTH: 492
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-72

```

```

Alignment Scores:
Pred. No.: 1,31 Length: 492
Score: 78.50 Matches: 34
Percent Similarity: 38.94% Conservative: 10
Best Local Similarity: 30.09% Mismatches: 52
Query Match: 5.27% Indels: 17
DB: 4 Gaps: 5

```

US-09-763-902B-10 (1-830) x US-09-413-814-72 (1-492)

```

Oy      309 CCAGGAGTCCAAATCAGCGTGTAGACTGCCAGCTGTTCATGATGAAAGCCAGACCA 368
Db      28 ProThrProArGProArGlyTyrPArGTrpThrSerPheThrSerArArGProSerAla 47
Oy      369 GCCAGACACAGG-----TCCTTGTTCATCTGTGATCAAGTGAAGTTTGAGGCGA 419
Db      48 SerArProAlaAlaLaseGlySerSerSerGlySerValArGAlaProSerProArGly 67
Oy      420 ACAAAACAGGAGCTTCAACCAAGACT-----TCATCTGACCCCGCCAGG 464
Db      68 AsnGlySerGlyCysSerThrThrThrLeuProAlaGluAlaLaserTrpProArGArG 87
Oy      465 CCT-----CACCCAGACACAGAGT-----GGAGATGCCAAGTACTGCTTCC 509
Db      88 ProCysArGmetProThrSerSerCysSerSerProSerArGserArGAlaLeuSer 107
Oy      510 GCTTCAGAGACTGGGCCAGCTAGTGG-----GTTGCAGAGGCTCTTCTGCT 557
Db      108 LeuHisLeuLeuProProSerSerGlyArGArGProSerGlyAsnIleAlaLeuSerAla 127
Oy      558 TCATTACAGCCCTACTCTGTAGAGAAATGCAAACTGGA 596
Db      128 AlaLeuSerProProAlaGlyProArGAlaLeuProArG 140

```

RESULT 8

```

US-09-579-181-2
Sequence 2, Application US/09579181
Patent No. 6365372
GENERAL INFORMATION:
APPLICANT: Chiviva, John
APPLICANT: Yaciuk, Peter
TITLE OF INVENTION: SMT2 Related CBP Activator Protein (SRCAP)
FILE REFERENCE: 16153-4247
CURRENT APPLICATION NUMBER: US/09/579,181
CURRENT FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: 60/136,620
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2972
TYPE: PRT
ORGANISM: Human
US-09-579-181-2

```

```

Alignment Scores:
Pred. No.: 3,38 Length: 2972
Score: 78.50 Matches: 55
Percent Similarity: 32.17% Conservative: 19
Best Local Similarity: 23.91% Mismatches: 88
Query Match: 5.27% Indels: 68
DB: 4 Gaps: 7

```

US-09-763-902B-10 (1-830) x US-09-579-181-2 (1-2972).

```

Oy      27 CTTCACCTTACTTCCCTGAGCCCTGTTCCCAAGCAGAGAAATACCTGCTGGA 86
Db      1221 LeuAlaProValValProAlaAlaProGlyProProSerLeuGln-----ProSerGly 1238
Oy      87 GCCCTCCTTCATAGAACCAAGAGATGCATCTGTGATTCAGACCTATGTGAGTACAG 146
Db      1239 Ala-----Ser 1240

```



CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Moessner, Warren D  
 REGISTRATION NUMBER: 30,440  
 REFERENCE/DOCKET NUMBER: 150,157US1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-359-3260  
 TELEFAX: 612-359-3263  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 480 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-570-227A-2

Alignment Scores:  
 Pred. No.: 1.49 Length: 480  
 Score: 78.00 Matches: 65  
 Percent Similarity: 33.57% Conservative: 29  
 Best Local Similarity: 23.21% Mismatches: 101  
 Query Match: 5.24% Indels: 85  
 DB: 2 Gaps: 14

US-09-763-902B-10 (1-830) x US-08-570-227A-2 (1-480)

```

OY 11 CATTACGCGGCTCTCTCAACCTTACTCC-----TCAGCCCGCTGGTCCCAAG 64
DB 118 Hispelyserleuserapthralalysprohislalalapro--Pheylsglu 136
OY 65 GCAGAGAAATACCTGTGGAGCCCTCTCCATGAAACAGATGSCATCTGGAT 124
DB 137 Glugluysserprovalserlarpolylsleuprolysalaglnlathrservalile 156
OY 125 TTCAGACCTATGTGATCAGCGCTGCAGA-----GCTGCTAGAGAG 166
DB 157 ARGHSTHIALASPALAGlnleucysasnhslnthrcysprometlyslalaser 176
OY 167 TTTCGATGTCTACTACACCATGATAGCGCGCGCTTG-----CTGTCC 217
DB 177 Ileleuasnlyrglnasnasnserpheargatrgatrhnlslenuasnvalglulalaia 196
OY 218 CGCGTGATACGGGACAGCCACCTGTCTGATGCAATGCT----- 262
DB 197 Arglyasnllserprocyalalavalserproasnargserlyscysgluargasnthr 216
OY 263 -----GTTTCAGACAGATCCTTGAGTGGATTTTGAATGTGCTCCAGC 313
DB 217 Valalaspvalasplulysalaserlalaileutyrgasphservalproser 236
OY 314 GAGTTCCAAATCAGCGGTAGACTGCCAGCTGTTCATGATGAGACACACAGCCAG 373
DB 237 Glu-----Thvalillecysargserglnproalaprovalserproindn 252
OY 374 ACCAGGCTCTGT----- 391
DB 253 Lysservalleuvalserproproalavalseralaglyvalproprokemp 272
OY 392 ATCTGTGATCAGTGAAGTTGAGGGAACAACAAGGAGCTTCAACAGACTTCATC 451
DB 273 Ilecygslmetvalproleuprolaasn-Asnprovalvalthrvalvalprose 292
OY 452 CTGACGCCCGCAG----- 464
DB 292 fththproserglnproproalavalcyserprovalvalphenetglynlnva 312
OY 465 -----CTTCACCCACACACAGTGTGGAAGTCCCA 496
  
```

```

DB 312 lProlysglyalavalmetphevalvalproglinprovalvalglinserlyserpropr 332
OY 497 AGTGACGCT-----TCCGCTTCAGACAGCGGCGAGCTGTGGGTGCAG 544
DB 332 ovalvalserproasnlythrargleuserprollaalaproalaproly----- 349
OY 545 AGCTCTCTTTGCTTCATTCAGCCCTAGCTGTAGAAATGCAACCTGAGTC----- 599
DB 350 -----Pheerproseralalaal--Lysvallnhproclnleasps 363
OY 600 -TCAGAGATGTGAGAACACAAATTCATTTCTGTG-----TCGAGACACT 647
DB 363 rserargyle-----Argserhisllecysserhlsproglycylslylystr 379
  
```

RESULT 11  
 US-09-077-991-2  
 Sequence 2, Application US/09077991  
 Patent No. 6207375  
 GENERAL INFORMATION:  
 APPLICANT: Sudramaniam, M.  
 APPLICANT: Spelsberg, T.C.  
 APPLICANT: Roche, P.C.  
 TITLE OF INVENTION: TGF-Beta Inducible early factor-1  
 FILE REFERENCE: (TGF-1) and a method to detect breast cancer  
 CURRENT APPLICATION NUMBER: US/09/077,991  
 EARLIER APPLICATION NUMBER: PCT/US96/19555  
 EARLIER FILING DATE: 1996-12-11  
 EARLIER APPLICATION NUMBER: US 08/570,227  
 EARLIER FILING DATE: 1995-12-11  
 NUMBER OF SEQ ID NOS: 13  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 2  
 LENGTH: 480  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-077-991-2

Alignment Scores:  
 Pred. No.: 1.49 Length: 480  
 Score: 78.00 Matches: 65  
 Percent Similarity: 33.57% Conservative: 29  
 Best Local Similarity: 23.21% Mismatches: 101  
 Query Match: 5.24% Indels: 85  
 DB: 4 Gaps: 14

US-09-763-902B-10 (1-830) x US-09-077-991-2 (1-480)

```

OY 11 CATTACGCGGCTCTCTCAACCTTACTCC-----TCAGCCCGCTGGTCCCAAG 64
DB 118 Hispelyserleuserapthralalysprohislalalapro--Pheylsglu 136
OY 65 GCAGAGAAATACCTGTGGAGCCCTCTCCATGAAACAGATGSCATCTGGAT 124
DB 137 Glugluysserprovalserlarpolylsleuprolysalaglnlathrservalile 156
OY 125 TTCAGACCTATGTGATCAGCGCTGCAGA-----GCTGCTAGAGAG 166
DB 157 ARGHSTHIALASPALAGlnleucysasnhslnthrcysprometlyslalaser 176
OY 167 TTTCGATGTCTACTACACCATGATAGCGCGCGCTTG-----CTGTCC 217
DB 177 Ileleuasnlyrglnasnasnserpheargatrgatrhnlslenuasnvalglulalaia 196
OY 218 CGCGTGATACGGGACAGCCACCTGTCTGATGCAATGCT----- 262
DB 197 Arglyasnllserprocyalalavalserproasnargserlyscysgluargasnthr 216
OY 263 -----GTTTCAGACAGATCCTTGAGTGGATTTTGAATGTGCTCCAGC 313
DB 217 Valalaspvalasplulysalaserlalaileutyrgasphservalproser 236
  
```

OY 314 GAGTTCGAATACAGCGGTAGACGACGACCTGTTCATGATGAGACCAACGACGACG 373  
 DB 237 GGU-----ThValIleCysArgSerGlnProAlaProValSerProGlnIn 252  
 OY 374 ACCAGGCTCTGTT-----GTC 391  
 DB 253 LysSerValIleValSerProProAlaValSerIleGlyValProProMetProVal 272  
 OY 392 ATCTGTGATCAGTCAAGTGTGAGGAGCAACACAGGAGCTTACACAGACTTCATC 451  
 DB 273 IleCysGlnMetValProLeuProAlaAsn-AsnProValValThrThrValValProSe 292  
 OY 452 CTGACCGCCGACG-----464  
 DB 292 rThProProSerGlnProProAlaValCysProProValValPheMetGlyThGlnVa 312  
 OY 465 -----CCTCACCCAGCAACAGACGTGAGAGATCGCA 496  
 DB 312 lProLysGlyAlaValMetPheValValProGlnProValValGlnSerSerLysProBr 332  
 OY 497 AGTCACTGCT-----TCGCTTCACAGACTGGGCGACCTAGTGGGGGTGGCAG 544  
 DB 332 oValValSerProAsnGlyThrArgLeuSerProIleAlaProAlaProGly-----349  
 OY 545 AGGTCTCTTCTTATTCAGCCCTAGCTCTGTAGGAATGCAAACTCGACTC-----599  
 DB 350 -----PheSerProSerIleAla--LysValThrProGlnIleLysPse 363  
 OY 600 -TCAGAGATGTAGAGAACAGATTTCATTTCTGTTGT-----TCGGGAGACACT 647  
 DB 363 rSerArgIle-----ArgSerHisIleCysSerHisProGlyCysGlyLysThr 379  
 RESULT 12  
 US-08-290-301-82  
 Sequence 82, Application US/08290301  
 Patent No. 5792921  
 GENERAL INFORMATION:  
 APPLICANT: Lonsborough, John  
 APPLICANT: Tunnela, Outi  
 APPLICANT: Palva, Tupio  
 APPLICANT: Holmstrom, Kjell-Ove  
 APPLICANT: Mandel, Bjorn  
 TITLE OF INVENTION: Increasing the trehalose content  
 TITLE OF INVENTION: of organisms by transforming them with combinations of  
 TITLE OF INVENTION: the structural genes for trehalose synthase.  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Alko Ltd.  
 STREET: PO Box 350  
 CITY: Helsinki  
 STATE: Finland  
 COUNTRY: Finland  
 ZIP: SF-00101  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
 COMPUTER: IBM PC/XT/AT  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: WP5.1 file exported as DOS text file  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/290, 301  
 FILING DATE: 15 August 1994  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FI 943133  
 FILING DATE: 29 June 1994  
 APPLICATION NUMBER: PCT/FI93/00049  
 FILING DATE: 15 February 1993  
 APPLICATION NUMBER: 07/841,997  
 FILING DATE: 28 February 1992  
 APPLICATION NUMBER: 07/836,021  
 FILING DATE: 14 February 1992  
 ATTORNEY/AGENT INFORMATION:

NAME: Kubovcik, Ronald J.  
 NAME: Lydon, James C.  
 REGISTRATION NUMBER: 25,401  
 REGISTRATION NUMBER: 30,082  
 REFERENCE/DOCKET NUMBER: LAIN-001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 467-6300  
 TELEFAX: (202) 466-2006  
 INFORMATION FOR SEQ ID NO: 82:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1098 amino acids  
 TYPE: Amino acid  
 TOPOLOGY: Linear  
 MOLECULE TYPE: Polypeptide  
 HYPOTHETICAL: Yes  
 US-08-290-301-82  
 Alignment Scores:  
 Pred. No.: 2,3 Length: 1098  
 Score: 78.00 Matches: 27  
 Percent Similarity: 47.37% Conservative: 9  
 Best Local Similarity: 35.53% Mismatches: 28  
 Query Match: 5.24% Indels: 12  
 Gaps: 3  
 US-09-763-902b-10 (1-830) x US-08-290-301-82 (1-1098)  
 OY 381 TCCTTTTTCATCCTGTGATCAGTGTGAGGGAACAAACAGGAGCTTCAACC 440  
 DB 186 SerLeuLeuAsnGlnThrSerGlnThrSerLeuGlyGlyProAsnGlnHisIleValThr 205  
 OY 441 AGAATTCATCCTCGAGCCGCGCCAGCTTCACCCAGCAACAGTGTGAGAGATCGCACTG 500  
 DB 206 ProLysSerArgIleGlyAsnArgPro-----ThrSerAlaAla 218  
 OY 501 ACTGCTTCGCTTCACAGACT-----GGCCACGCTAGTGGGGGTGGCAGAGTCTCTT 554  
 DB 219 ThrSerLeuValAsnArgThrLysGlnGlySerAlaSerSerLysSerGly-----236  
 OY 555 GCTTCATTCAGCCCTAGCTCTGTAGGAATGCAAACTCGACTCTCA 602  
 DB 237 ---SerSerAlaProProSerIleLysArgIleThrProHisLeuThr 251  
 RESULT 13  
 US-09-013-598-82  
 Sequence 82, Application US/09013598  
 Patent No. 6323001  
 GENERAL INFORMATION:  
 APPLICANT: Lonsborough, John  
 APPLICANT: Tunnela, Outi  
 APPLICANT: Palva, Tupio  
 APPLICANT: Holmstrom, Kjell-Ove  
 APPLICANT: Mandel, Bjorn  
 TITLE OF INVENTION: Increasing the trehalose content  
 TITLE OF INVENTION: of organisms by transforming them with combinations of  
 TITLE OF INVENTION: the structural genes for trehalose synthase.  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Alko Ltd.  
 STREET: PO Box 350  
 CITY: Helsinki  
 STATE: Finland  
 COUNTRY: Finland  
 ZIP: SF-00101  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
 COMPUTER: IBM PC/XT/AT  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: WP5.1 file exported as DOS text file  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/013, 598  
 FILING DATE:



CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/290,301  
FILING DATE:  
APPLICATION NUMBER: PCT/E193/00049  
FILING DATE: 15 February 1993  
APPLICATION NUMBER: 07/841,997  
FILING DATE: 28 February 1992  
APPLICATION NUMBER: 07/836,021  
FILING DATE: 14 February 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kubovcik, Ronald J.  
REGISTRATION NUMBER: 25,401  
REFERENCE/DOCKET NUMBER: 30,082  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 466-6300  
TELEFAX: (202) 466-2006  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1098 amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: Polypeptide  
HYPOHETICAL: Yes  
US-09-013-598-82

Alignment Scores:  
Pred. No.: 2,3  
Score: 78.00  
Percent Similarity: 47.37%  
Best Local Similarity: 35.53%  
Query Match: 5.24%  
DB: 4  
Length: 1098  
Matches: 27  
Conservative: 9  
Mismatch: 28  
Indels: 12  
Gaps: 3

US-09-763-902b-10 (1-830) x US-09-013-598-82 (1-1098)

QY 381 TCGTGTCTGTCATGTCGTGATGATGAGTGGAGGAGCAACAGCGGACTTCACACC 440  
DB 186 SerLeuLeuAlaAsnThrSerLInThrSerLeuGlnGlyProAsnAsnHisLeuAlaThr 205  
QY 441 AGAATTCATCTGAGTCCGCCGCCAGGCTCAGCCAGACACAGTGTGAAGATGCGAGAG 500  
DB 206 ProLysSerArgAlaGlyAsnArgPro-----ThSerAlaAla 218  
QY 501 ACTGCTTCGCTCCAGACT-----GGCCAGCTAGTGGGGGTGGCAGAGGCTCTTT 554  
DB 219 ThrSerLeuValAsnArgThrLysGlnGlySerAlaSerSerGlySerGly----- 236  
QY 555 GCTTCATTCAGCCCTGAGTGTGAGAGAAATGCAAACTGCACTCTCA 602  
DB 227 ---SerSerAlaProProSerLeuLysArgGlyLeuThrProHisLeuThr 251

RESULT 14  
US-08-216-260-4  
Sequence 4, Application US/08216260  
Patent No. 5837261  
GENERAL INFORMATION:  
APPLICANT: Ingalls, Stephen C.  
APPLICANT: Bousnell, Michael E. G.  
APPLICANT: Minson, Anthony C.  
TITLE OF INVENTION: VIRAL VACCINES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohnbach, Teet, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/216,260  
FILING DATE: 21-MAR-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9305710.7  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9324964.7  
FILING DATE: 06-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/030,073  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/168,643  
FILING DATE: 16-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreyer, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-59886/WHD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 838 amino acids  
TYPE: amino acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
US-08-216-260-4

Alignment Scores:  
Pred. No.: 2,64  
Score: 77.00  
Percent Similarity: 35.45%  
Best Local Similarity: 25.45%  
Query Match: 5.17%  
DB: 2  
Length: 838  
Matches: 56  
Conservative: 22  
Mismatch: 66  
Indels: 76  
Gaps: 13

US-09-763-902b-10 (1-830) x US-08-216-260-4 (1-838)

QY 51 CCGTGTCCCAAGCAGAGAGAAATACCTGCTGAGCCCTTCATAGACAGAGA 110  
DB 53 ProAsnThrProAspProGlnLysProProArgGlyPheLeuAlaPro----- 68  
QY 111 TGGCATCTGTGATTTCAAGACCTATGTGATCAGGCCCTGAGAGCTGAGAGATTG 170  
DB 69 -----ProAspGlnLeu-----AsnLeu 74  
QY 171 TCAATGCTACTACACCA-----CCATGATTAACGGCGGCTTGCTGCCGCTGT 224  
DB 75 ThrThrAlaSerLeuProLeuLeuArgTrpGlnGluArgPheCysPheValLeuVal 94  
QY 225 ACATGGCAGACACCCCTGCTGTGAGATGCAATGCTGTTTCAGCAAGAAAT----- 278  
DB 95 ThrThrAlaGlnPheProArgAspProGlyGlnLeuLeuTrpIleProLysThrTyLeu 114  
QY 278 ----- 278  
DB 115 LeuGlnArgProProAsnAlaSerLeuProAlaProThrThrValGluProThrAlaGln 134  
QY 279 -----CCTTGAAGATGTTTGAATGTTGCTT-----CCAGGAATTC----- 320  
DB 135 ProProProSerValAlaProLeuLysGlyLeuLeuHisAsnProAlaAlaSerValLeu 154  
QY 321 -----AAATCAGCGGTGAGACTGCCAGCTGTTGATGATGAAGCAGACAGCAGCA 374  
DB 155 LeuArgSerArgAlaTrpValThr-----PheSerAlaValAlaProAsp----- 168



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 19:31:14 ; Search time 31.503 Seconds  
(without alignments)  
5701.778 Million cell updates/sec

Title: US-09-763-902B-10

Perfect score: 1489  
Sequence: 1 tgaacttgcatcactgctg.....ataattaataacacatg 830

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Zgapop 6.0, Zgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 835558

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus\_n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO/US09/63902/runat\_24062003\_130343\_12522/app\_query.fasta.1.2254  
-DB=PublishedApplications\_AA -QFM=FASTA -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09763902.ecgn.1.1.24@runat\_24062003\_130343\_12522  
-NCPU=6 -ICPU=3 -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONLOG -DEV=TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: PublishedApplications\_AA.\*

1: /cgn2.6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2.6/ptodata/2/pubpaa/FCI\_NEW\_PUB.pep.\*  
3: /cgn2.6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2.6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2.6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2.6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2.6/ptodata/2/pubpaa/FCIS\_PUBCOMB.pep.\*  
8: /cgn2.6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2.6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2.6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2.6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2.6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2.6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2.6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	813	54.6	164	US-09-764-864-1181
2	579	38.9	10	Sequence 1181, Ap
3	115.5	7.8	449	US-09-764-864-1180
4	105.5	7.1	9	Sequence 272, App
				Sequence 318, App

	5	105.5	7.1	477	10	US-09-925-301-977	Sequence 977, App
6	89	6.0	326	10	US-09-764-864-1044	Sequence 1044, Ap	
7	86.5	5.8	462	10	US-09-814-777A-99	Sequence 99, Appl	
8	85.5	5.7	802	10	US-09-907-495-2	Sequence 2, Appl1	
9	85.5	5.7	1604	10	US-09-746-491-8	Sequence 8, Appl1	
10	83.5	5.6	197	9	US-10-149-328-8	Sequence 8, Appl1	
11	83.5	5.6	475	10	US-09-741-669-458	Sequence 458, App	
12	80.5	5.4	831	9	US-10-273-680-2	Sequence 2, Appl1	
13	80	5.3	1276	10	US-09-982-610-24	Sequence 24, Appl1	
14	78.5	5.3	725	9	US-10-160-501-17	Sequence 17, Appl1	
15	78.5	5.3	2971	9	US-10-146-472-50	Sequence 50, Appl1	
16	77.5	5.2	221	9	US-10-102-806-603	Sequence 603, Appl1	
17	77.5	5.2	1057	9	US-10-189-971-6	Sequence 6, Appl1	
18	77.5	5.2	1251	9	US-10-189-971-16	Sequence 16, Appl1	
19	77.5	5.2	1342	9	US-10-189-971-24	Sequence 24, Appl1	
20	77.5	5.2	1512	9	US-10-189-971-10	Sequence 10, Appl1	
21	77.5	5.2	1570	9	US-10-189-971-12	Sequence 12, Appl1	
22	77.5	5.2	1628	9	US-10-189-971-1	Sequence 2, Appl1	
23	76	5.1	190	9	US-09-738-626-6057	Sequence 6057, Ap	
24	76	5.1	398	9	US-10-097-340-248	Sequence 248, App	
25	76	5.1	1106	9	US-10-157-031-30	Sequence 30, Appl1	
26	76	5.1	2530	9	US-10-184-644-155	Sequence 155, App	
27	76	5.1	2530	9	US-10-184-634-155	Sequence 155, App	
28	75.5	5.1	242	10	US-09-764-853-609	Sequence 609, App	
29	75.5	5.1	242	10	US-09-764-898-199	Sequence 199, App	
30	75.5	5.1	619	9	US-09-976-059-30	Sequence 30, Appl1	
31	75.5	5.1	785	10	US-09-801-368-348	Sequence 348, App	
32	75.5	5.1	963	9	US-10-140-164-2	Sequence 2, Appl1	
33	75.5	5.1	1027	9	US-10-140-164-4	Sequence 4, Appl1	
34	75.5	5.1	2871	9	US-10-146-473-41	Sequence 41, Appl1	
35	75	5.0	316	9	US-10-153-668-134	Sequence 134, App	
36	75	5.0	323	9	US-10-153-668-132	Sequence 132, App	
37	75	5.0	370	9	US-10-153-668-130	Sequence 130, App	
38	75	5.0	410	9	US-09-870-759-48	Sequence 48, Appl1	
39	75	5.0	473	9	US-09-978-292A-400	Sequence 400, App	
40	75	5.0	473	9	US-09-978-697A-400	Sequence 400, App	
41	75	5.0	473	9	US-09-978-192A-400	Sequence 400, App	
42	75	5.0	473	9	US-09-999-832A-400	Sequence 400, App	
43	75	5.0	473	9	US-09-978-189-400	Sequence 400, App	
44	75	5.0	473	9	US-10-028-072-382	Sequence 382, App	
45	75	5.0	473	9	US-10-121-049-382	Sequence 382, App	

#### ALIGNMENTS

RESULT 1  
US-09-764-864-1181  
Sequence 1181, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764,864  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1181  
LENGTH: 164  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (2)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (3)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (5)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE



```

DB 34 TYRGLYARGANSERSETYRVAHISGLYVALASPALASERGLYLSERPROGLINGU 53
QY 260 GCTGTTTACGACAGAAATCCCTGAGTGTGTTTGAATGTCCTCCAGCGACTC 319
DB 54 AIAVALTYRGLYGLNASHNAPRLEHISHSLYSVALLEUSERLEUASHNAPRHESEGLUCYS 73
QY 320 CAATACAGCGGTGAGTCCAGCCCTGTCATGATGAAGCAACCAAGCCAGACGACG 379
DB 74 HISHRYSLE-----ARGHISVALASPALAHSHSLATHTLEUSERSP---GLY 89
QY 380 GTCTTTGTCATCTGTGATCATGATGAAGTTGAGGGAACAAACAGGACTTCAAC 439
DB 90 VALVALVALGlnValMetGlyLeuLeuSerAsnSerGlyGlnProGlnArgLysPheMet 109
QY 440 CAGAACTTCATCCAGACCCGAGCCCTCACCAGACACACAGTGTGAAGTCCGACT 499
DB 110 GlnThrPheValLeuAlaProGlnLysSerValProAsnLysPheTyr---ValHisAsn 128
QY 500 GACTGCTTCGCTTCAGGAC 520
DB 129 ASPMETPHEARGTYRGLNASP 135

```

## RESULT 4

```

US-09-919-039-318
; Sequence 318, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919, 039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 318
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 2477616CD1
US-09-919-039-318

```

## Alignment Scores:

```

Pred. No.: 0.0317 Length: 466
Score: 105.50 Matches: 36
Percent Similarity: 44.88% Conservative: 21
Best Local Similarity: 28.35% Mismatches: 57
Query Match: 7.09% Indels: 13
DB: 9 Gaps: 4

```

US-09-763-902b-10 (1-830) x US-09-919-039-318 (1-466)

```

QY 164 GAGTTTGTCAATGTCTACTACACCAACGATGAAGCGGGCGTTTGTCTCCGCGCTG 223
DB 14 GlnPheValArgLysThrLeuLeuAsnGlnAlaProAspMetLeuHisArgPhe 33
QY 224 TACATGGGACACCCCTGCTGTGATGGC-----AAT 259
DB 34 TYRGLYLYSASNSERSETYRVAHISGLYGLYLEUASPSERASNGLYLSERPROALASP 53
QY 260 GCGTTTCAAGACAAATCCCTGAGTGTGTTTGAATGTTGCTTCAGGAGACTC 319
DB 54 AIAVALTYRGLYGLNASHNAPRLEHISHSLYSVALLEUSERGLNASPHE 70
QY 320 CAATACAGCGGTGAGTCCAGCCCTGTCATGATGAAGCAACCAAGCCAGACGACG 379
DB 71 THASNCYSHISHRYSLEARGHISVALASPALAHSHSLATHTLEUASNAPR---GLY 89
QY 380 GTCTTTGTCATCTGTGATCATGATGAAGTTGAGGGAACAAACAGGACTTCAAC 439

```

```

DB 90 VALVALVALGlnValMetGlyLeuLeuSerAsnAsnGlnAlaLeuArgArgPheMet 109
QY 440 CAGAACTTCATCCAGACCCGAGCCCTCACCAGACACACAGTGTGAAGTCCGACT 499
DB 110 GlnThrPheValLeuAlaProGlnLysSerValAlaHisLysPheTyr---ValHisAsn 128
QY 500 GACTGCTTCGCTTCAGGAC 520
DB 129 ASPMETPHEARGTYRGLNASP 135

```

## RESULT 5

```

US-09-925-301-977
; Sequence 977, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 977
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (471)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (473)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-977

```

## Alignment Scores:

```

Pred. No.: 0.032 Length: 477
Score: 105.50 Matches: 36
Percent Similarity: 44.88% Conservative: 21
Best Local Similarity: 28.35% Mismatches: 57
Query Match: 7.09% Indels: 13
DB: 10 Gaps: 4

```

US-09-763-902b-10 (1-830) x US-09-925-301-977 (1-477)

```

QY 164 GAGTTTGTCAATGTCTACTACACCAACGATGAAGCGGGCGTTTGTCTCCGCGCTG 223
DB 25 GlnPheValArgLysThrLeuLeuAsnGlnAlaProAspMetLeuHisArgPhe 44
QY 224 TACATGGGACACCCCTGCTGTGATGGC-----AAT 259
DB 45 TYRGLYLYSASNSERSETYRVAHISGLYGLYLEUASPSERASNGLYLSERPROALASP 64
QY 260 GCTGTTTCAAGACAAATCCCTGAGTGTGTTTGAATGTTGCTTCAGGAGACTC 319
DB 65 AIAVALTYRGLYGLNASHNAPRLEHISHSLYSVALLEUSERGLNASPHE 81
QY 320 CAATACAGCGGTGAGTCCAGCCCTGTCATGATGAAGCAACCAAGCCAGACGACG 379
DB 82 THASNCYSHISHRYSLEARGHISVALASPALAHSHSLATHTLEUASNAPR---GLY 100
QY 380 GTCTTTGTCATCTGTGATCATGATGAAGTTGAGGGAACAAACAGGACTTCAAC 439
DB 101 VALVALVALGlnValMetGlyLeuLeuSerAsnAsnGlnAlaLeuArgArgPheMet 120
QY 440 CAGAACTTCATCCAGACCCGAGCCCTCACCAGACACACAGTGTGAAGTCCGACT 499
DB 121 GlnThrPheValLeuAlaProGlnLysSerValAlaHisLysPheTyr---ValHisAsn 139

```



Db	296	AlaProValProArgGlyThrGlyGlnThrGlyProAlaArgHisValLeuProArgGly	315																								
QY	471	-----CCAGCACACAGTGT-----GGAGATGGCAAGTACTGCTCCCGCTCC	515																								
Db	316	GlunIInProAspSerArgGlyThrProAlaAlaArgSerIleThrAlaArgAlaSer	335																								
QY	516	AGAGCT-----GGGCCAGCTGATGGGGGTGGCAGAGCT-----	548																								
Db	336	ArgAlaAlaAlaProProGlyProCysSerAlaSerSerArgSerProArgAspArgSer	355																								
QY	549	-----CTCTTGCTTCATTCAGCCCTAGC	572																								
Db	356	GluproAlaValArgSerAlaLeuSerThrAlaCysMetPheGlySerMetSerGlnPro	375																								
QY	573	TCGTAGAGAAATGCMAACCTCGACTCTCAAGAGATGTAGAGACACAAAGTTATTCTGT	632																								
Db	376	ProArgSerGlnCysArgPro-----	382																								
QY	633	TGTGTCGGAGACATGCGAGACTCCAGCTGGCGAGGTGAATCTTTTGTGTGGCAAG	692																								
Db	383	CysAlaArgSerThrSerGlnAlaThrLeuProGlyLeu-----	395																								
QY	693	TTTAGAGAGTCCCTTCTCGATATATATACTGTGTGCA-----	731																								
Db	396	-----LeuAlaThrCysLeuGlyGlyProLeuArgGlyCysLeu	408																								
QY	732	TAGTTTCCT-----	740																								
Db	409	GlupheProArgValProGlyLeupheGlnGlnAlaArgAlaGlnAspLeuAlaGln	428																								
QY	741	-----TTCAAGTACTATAA	755																								
Db	429	LeuproGlyLeuHisPheSerThrCysSerPheSerCysSerValPheSerThrThrArg	448																								
QY	756	CTTTCTCATTTTCTTACTTGCCTCA	779																								
Db	449	LeuThrTyrPheLeuLeuLeuCysPro	456																								
<p>RESULT 8  US-09-907-495-2  Sequence 2, Application US/09907495  Patent No. US20020081696A1  GENERAL INFORMATION:  APPLICANT: Meyers, Rachel  TITLE OF INVENTION: 32529, A NOVEL HUMAN GANINE NUCLEOTIDE EXCHANGE FACTOR  TITLE OF INVENTION: FAMILY MEMBER AND USES THEREOF  FILE REFERENCE: MNI-175  CURRENT APPLICATION NUMBER: US/09/907,495  PRIOR FILING DATE: 2001-07-16  PRIOR APPLICATION NUMBER: 60/218,383  PRIOR FILING DATE: 2000-07-14  NUMBER OF SEQ ID NOS: 3  SOFTWARE: PatentIn Ver. 2.0  SEQ ID NO 2  LENGTH: 802  TYPE: PRT  ORGANISM: Homo sapiens  US-09-907-495-2</p>																											
<p>Alignment Scores:</p> <table border="0"> <tr> <td>Pred. No.:</td><td>5.11</td><td>Length:</td><td>802</td></tr> <tr> <td>Score:</td><td>85.50</td><td>Matches:</td><td>64</td></tr> <tr> <td>Percent Similarity:</td><td>35.00%</td><td>Conservative:</td><td>27</td></tr> <tr> <td>Best Local Similarity:</td><td>24.62%</td><td>Mismatches:</td><td>98</td></tr> <tr> <td>Query Match:</td><td>5.74%</td><td>Indels:</td><td>71</td></tr> <tr> <td>DB:</td><td>10</td><td>Gaps:</td><td>14</td></tr> </table>				Pred. No.:	5.11	Length:	802	Score:	85.50	Matches:	64	Percent Similarity:	35.00%	Conservative:	27	Best Local Similarity:	24.62%	Mismatches:	98	Query Match:	5.74%	Indels:	71	DB:	10	Gaps:	14
Pred. No.:	5.11	Length:	802																								
Score:	85.50	Matches:	64																								
Percent Similarity:	35.00%	Conservative:	27																								
Best Local Similarity:	24.62%	Mismatches:	98																								
Query Match:	5.74%	Indels:	71																								
DB:	10	Gaps:	14																								
<p>US-09-763-902B-10 (1-830) x US-09-907-495-2 (1-802)</p>																											
QY	24	TCCTCTCAACCTTACTCTCTCGAGCC-----CTGTGTTCCCAAGGCAAGG	71																								
Db	39	AlaLeuLeuProSerProValCysLeuAspLeuPheProValAlaProGlnGlnLeu	58																								

OY	72	AAATACCCCTGATGAGACCCCTCCCTTCATGAAACCAAGATGCACTGTGATTTCAAGA	131
Db	59	ArgAlaProGly-----SerArgTyrSerLeuGlyThrProAla	71
OY	132	CGTATGTGATGACGCGCTGCAGACTGCTGCAGAGAGTTTGTCAATGCTCTACTACACCACA	191
Db	72	ProLeu-----GlnGlyLeuLeuTyrProLeuSerProGlyGlySerAspThr	87
OY	192	TGGATA---AGCGGCGCGCTT-----TGGTGTCCCGCTGTACATGG	230
Db	88	GlnIleHisSerGlyLeuMetAlaGlyProSerArgAlaGlySerTyrProHisCysProGly	107
OY	231	GCACAGACCA-----CCCTGGTGTGAAATGCAATGCTGTTTCAGGACAAAG	275
Db	108	AlaGlnProArgAlaLeuGlnGlyProTyrSerProAlaHisThrGlnProGln-----	125
OY	276	AATGCTTGAAGTGAATTTTGAATGTGCTGCTTCAGCGCAAGTTCCAAATCAGCGTGTGG	335
Db	126	-----ArgArgAlaSerHisGlySerGlnIleCysSerAlaThrArg	139
OY	336	ACTGCCAGCGCTGTTCC-----ATG	353
Db	140	LysMetArgValTyrGlnArgGlnGluValProGlyCysArgProGlnAlaHisAlaValPhe	159
OY	354	ATGAGCGCACACCAAGCC-----AGACCAGCGCTCTGG	386
Db	160	LeuIleProGlyGlnValValGlnGlnGlnAlaLeuSerThrGlnLeuProArgValGln	179
OY	387	TTTGTCATCTGTGATGATGATGATTTGAGGGGAACAACAACAGGACTTCAACAGAACT	446
Db	180	LeuSerGlySerThrArgValSerLeuGlnGlyProGlnArgArgPheSer---Ala	198
OY	447	TCATCTGACCGCCCGCCAGCGCTCACCCAGCAACACAGTGTGAAGATGCGCAAGTACTGT	506
Db	199	SerIleuMetThrArgLeuHisSerSerLeuArgLeuGlyIleArgAsnSer---AlaAla	217
OY	507	TCCGCTTCAGAGCTGGCGCCAGCTAGTGGGGGTGGCAGA---GGTCTCTTGGCTTCATTC	563
Db	218	ArgAlaLeuIleSerGlySerGlyThrGlyAlaAlaArgGlnGlnGlyValSerGlyMet	237
OY	564	AGCCCTAGCTCTGTAGAGAAATGCAAACTCGACTTCAGAGATGTGAGGACAAAGTT	623
Db	238	GlnAlaArgSerValGlnMetSerGlyAspArgValSerArgPro-----	252
OY	624	CATTCTGCTTTCGCGAGACACT-----GCAGACTCCACTGCGCGGAGTGTGAATCT	677
Db	253	-----AlaProGlyAspSerArgGlnGlyAspTyrSerGlnProArgLeuAspThr	269

RESULT 9

US-09-746-491-8

Sequence 8, Application US/09746491

Patent No. US20020137202A1

GENERAL INFORMATION:

APPLICANT: Burgess, Catherine E.

TITLE OF INVENTION: NO. US20020137202A1el Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 15966-621

CURRENT FILING DATE: 2000-12-20

PRIOR APPLICATION NUMBER: US5N 60/171,329

PRIOR FILING DATE: 1999-12-21

NUMBER OF SEQ ID NOS: 72

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 8

LENGTH: 1604

TYPE: PRT

ORGANISM: Homo.sapiens

US-09-746-491-8

Alignment Scores:

Pred. No.:	6-68	Length:	1604
Score:	85-50	Matches:	64
Percent Similarity:	35.00%	Conservative:	27

Best Local Similarity: 24.628  
Query Match: 5.748  
DB: 10  
Matches: 98  
Indels: 71  
Gaps: 14

US-09-763-902B-10 (1-830) x US-09-746-491-8 (1-1604)

```
OY 24 TCCTTCAACCTTACTTCCCTCAGCC-----CCGCGTTCGCCAGCAGAG 71
DB 841 AAlaleuIsProPserProValyIsleuAspLeuPheProValAlaProIuIndeu 860
OY 72 AAATACCTGGTGGAGCCCTCCCTCCATAGAACACAGATGCTGATTCAGAGA 131
DB 861 ArgAlaPProGly-----SerArgTyrSerLeuIndyThrProAla 873
OY 132 CCTATGTGATCAGCCTGCAGAGCTGCTGAGAGTGTTCATATGCTATACACCACCA 191
DB 874 ProIeu-----GlnGlyLeuIndyThrProLeuSerProGlyGlySerAspThr 889
OY 192 TGGATA---AGCGGGGGGTT-----TGCTGCCCGCCCTGTACATGG 230
DB 880 GluIleThrSerGlyGlyMetArgProSerArgAlaGlySerTyrProHisGlyProGly 909
OY 231 GCACAGCCA-----CCCTGCTGCTGGAATGCAATGCTGTTTCAGACAG 275
DB 910 AlAGInProProAlaLeuGluGlyProTyrSerProAlaGlnHisThrGlnProGln----- 927
OY 276 AATCCTTGAAGTGAATTTTGAATGTTCCCTCCAGGAGACTCCAAATCAGCGGTGAG 335
DB 928 -----ArgArgAlaSerHisGlySerGlyGlySerAlaIleThrArg 941
OY 336 ACTGCCACCTGTTC-----ATG 353
DB 942 LysMetArgValIleGlnArgGluGluValProGlyCysProGluAlaHisAlaValPhe 961
OY 354 ATGAAGCCACACACAGCC-----AGACCAAGGCTCTTG 386
DB 962 LeuGluProGlyGlnValIleGlnGluAlaLeuSerThrGlnGluProArgValGlu 981
OY 387 TTTCATCTGTGGATCAGTGAAGTTTGAAGGGAACAAACAGGAGCTTCAACAGAACT 446
DB 982 LeuSerGlySerThrArgValSerLeuGluGlyProGluArgArgPheSer---Ala 1000
OY 447 TCATCCTGACGCCAGCCCTCACCCACACAGAGTGTGAAGATCCCAAGTACTGCT 506
DB 1001 SerGluLeuMetThrArgLeuHisSerSerLeuArgLeuGlyArgAsnSer---AlaAla 1019
OY 507 TCCCTTCCAGAGTGGGCGCAGTAGTGGGGTGGCGA---GGTCTCTTTCCTCATTC 563
DB 1020 ArgAlaLeuIleSerGlySerGlyThrGlyAlaAlaArgGluGlyLysAlaSerGlyMet 1039
OY 564 AGCCCTACCTCTGTAGAAATGCAACCTGACTCTCAGATGATGTAGAAACAGATT 623
DB 1040 GlnAlaArgSerValGlnMetSerGlyAspArgValSerArgPro----- 1054
OY 624 CATTTCTGTTGTTCCGAGACACT-----GCAGACTCCACTGTCGGAGTTGACTCT 677
DB 1055 -----AlaProGlyAspSerArgGluGlyAspTyrSerGluProArgLeuAspThr 1071

RESULT 10
US-10-149-326-8
Sequence 8, Application US/10149326
Publication No. US20030100492A1
GENERAL INFORMATION:
APPLICANT: YAYON, AVNER
TITLE OF INVENTION: PROTOGLYCANS AND PHARMACEUTICAL COMPOSITIONS COMPRISING THEM
FILE REFERENCE: 01/22063
CURRENT APPLICATION NUMBER: US/10/149,326
CURRENT FILING DATE: 2002-06-05
PRIORITY APPLICATION NUMBER: IL 133318
PRIORITY FILING DATE: 1999-12-05
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.1
SEQ ID NO 8
```

LENGTH: 197  
TYPE: PRT  
ORGANISM: Gallus gallus  
US-10-149-326-8

Alignment Scores:  
Pred. No.: 4.84  
Score: 83.50  
Percent Similarity: 34.78%  
Best Local Similarity: 27.95%  
Query Match: 5.618  
DB: 9  
Length: 197  
Matches: 45  
Conservative: 11  
Mismatch: 45  
Indels: 60  
Gaps: 8

US-09-763-902B-10 (1-830) x US-10-149-326-8 (1-197)

```
OY 188 ACCATGATTAAGCGGGCGCTTGTCTGTCGCCCTGTACATGGCCACACCCCTGATC 247
DB 27 ThrMetAspAlaArg----- 31
OY 248 TGGATGCAATGCTGTTCAGA-----CAAGATCCTGAGTGTTTT 295
DB 32 ThrLeuAspAsnValGlySerGlyAspLeuProAspAspGluAspIleGlyIleThr 51
OY 296 GAATGTTGCTTCCAGGAGTCCCAATC----- 325
DB 52 ProHisLeuThrSerAspGluPheAspIleAspAspThrSerGlySerGlyAspTyrSer 71
OY 326 -----AGCGTGTAGACTGCCACCTGTTCAATGATAGACC 361
DB 72 AspTyrAspAspAlaIleTyrLeuThrThrValAspThrProAlaIleSerAspAsnTyr 91
OY 362 ACACCAACCCAGACCCGCTCTGTCTGATCAGTGAAGTTGAGGAGAAC 421
DB 92 IleProGlyAspThr-----GluArgLysMetGluGly 103
OY 422 AAACACGGGACTTACACAGAACTTACTGACCGGCCAGGCTCACC----- 472
DB 104 LysLysAsnThrMetLeuAspAsnGluIleIleProAspLysAlaSerProValGluAla 123
OY 473 -----AGCACACAGTGTGAAGATCCCAAGTACTCTCCGCTTC 515
DB 124 AsnLeuSerAsnLysIle---SerMetAlaSer---ThrIleAsnSerIlePheGluArg 142
OY 516 -----AGACTGGGCGCAGCTAGTGGGGTGGCAGAGT---CTCTTCTCTACTT 562
DB 142 GThrGluValLeuThrAlaLeuIleAlaGlyLysAlaValGlyLeuLeuPheAlaValPhe 162
OY 563 C 563
DB 162 e 162

RESULT 11
US-09-741-669-458
Sequence 458, Application US/09741669
Patent No. US20020022718A1
GENERAL INFORMATION:
APPLICANT: Forsyth, R. Allen
APPLICANT: Ohlsen, Karl L.
TITLE OF INVENTION: Genes identified as required for
FILE REFERENCE: ELITRA 009A
CURRENT APPLICATION NUMBER: US/09/741,669
CURRENT FILING DATE: 2000-12-19
PRIORITY APPLICATION NUMBER: US 60/173005
PRIORITY FILING DATE: 1999-12-23
NUMBER OF SEQ ID NOS: 481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 458
LENGTH: 475
TYPE: PRT
ORGANISM: Escherichia coli
US-09-741-669-458
```



## Alignment Scores:

Pred. No.:	6.8	Length:	475
Score:	83.50	Matches:	49
Percent Similarity:	40.20%	Conservative:	33
Best Local Similarity:	24.02%	Mismatches:	49
Query Match:	5.61%	Indels:	73
		Gaps:	10

US-09-763-902b-10 (1-830) x US-09-741-669-458 (1-475)

```

QY 62 AAGCAGAGAAATACCTGGTGGAGCCCTTCATGAGACAGAGATGGCATGTGTG 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 Glnserglusermetprrtprtysercylprrthleuaglvalleuaglthrv 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 122 GATTCAAGACCTTGTGGAT---CAGGCTGCAGAGCTGTGAGAGATTGTCAATG 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 Glutlelglarvalvalaspalaglprmetarppheprovalglntyrvalasn 254
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 179 TACTACACACCATGATGATAGAGGGGGCTTGGCTGCTCCGCTGACATGGGACAGCC 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 -----AigProasnleuaspphearglytyralaglythr--- 266
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 239 ACCCTGTGTGGAATGGCAATGCTGTTCAGAGACAGAAATCCTTGAGTGTGTTGAA 298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 267 -----LeuAlaserglyargvalgluvalaglylharvalys 279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 299 ATGTGCTTCC----- 310
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 Valleuprosarglyvalgluserasnvalalargilevalthrphaspolyasparg 299
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 311 -----AGGAGTTCACATTCAGC 328
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 300 GlutlualaphleaglgluAlaletherleuvalleuthraspdluleaspiliser 319
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 329 GTGTGAGATGCCAGCTGTTCATGATGAGCCACACAGAG---CAGACAGAGCTT 385
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 320 ArgglyaspleuleleuAlalaspdluleproAlalaserlaserAl 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 386 GTTGTATCTGTGATCAGTGA-----G 409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 Aspalval--TrrpetaalagluginProleuserProglylserTyrAspIleTysI 359
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410 TTGAGGGGAGACAAAC-----ACGGACTTCACCAACTCATCTGAGC 457
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 lealagllylsythrargalargvalaspolyileargtyrlnval-----AspI 377
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 458 GCCCAGGCTCACCAGACAGACAGATGTGAGATGCGAGAGTCTTCGCTTCAG 517
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 377 leasnasnleuthrlnarg--GluValGluasn-----LeuProleuasnG 392
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 518 GACTGGC 525
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 lyilegly 394
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

## RESULT 12

```

US-10-273-680-2
; Sequence 2, Application US/10273680
; Publication No. US20030099996A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Tomlinson, James E.
; TITLE OF INVENTION: VELPZ, A VASCULAR ENDOTHELIAL CELL
; TITLE OF INVENTION: SPECIFIC AND LIM DOMAIN CONTAINING MOLECULE AND USES
; FILE REFERENCE: MP101-317P1BM
; CURRENT APPLICATION NUMBER: US-10/273,680
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US 60/329,756
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 2
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-273-680-2

```

## Alignment Scores:

Pred. No.:	17.5	Length:	831
Score:	80.50	Matches:	59
Percent Similarity:	32.08%	Conservative:	26
Best Local Similarity:	22.26%	Mismatches:	87
Query Match:	5.41%	Indels:	93
		Gaps:	13

US-09-763-902b-10 (1-830) x US-10-273-680-2 (1-831)

```

QY 83 TGGAGCCCTCCTTCATGAAACAGATGGCATGTGATTTCAACCATATGTGAT 142
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 35 TrpValProProglyLeuArgProGluInleuTyrPhe----- 49
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 143 CAGGCTGCAGAGCTGAGGAGTTTGTCAATGCTACTACACCATGATAGCGG 202
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 50 ---Alacysleuprogluglulysval-----ProtyrValasnserProglylulys 66
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 203 CGGCTGTGTCTCCGCTGTACATGGCAGACAGCCCTGTGTGAAATGCCAATGCT 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67 HlsarglylelsglnleuLeuTyrGln-----LeuProProHlsaspsnglu 82
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 263 GTTTCAGACAGAAATCCTTGAGTGTGATTTTGAATGTTCCCTCCAGCAGTTCCAA 322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 83 ValArgTyrCysGlnserleuserGlu-----GluGlnulysGln 96
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 323 ATCAGCGTGTGACTGGCAGGCTGTTCATGATGAGCCACACAGCAGCAGC----- 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 LeuGlnValPheSerAlaGlnArg--LysLysGlnAlaLeuGlyArgGlyThrIleLys 115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 377 -----ACGTCCTGTGTTCATCTGTGATCAGTAAGTAATTTGAGGAGAA 424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116 LeuLeuserArgAlaValMethIshAlaValCys-GluGlnCysGlyLeuLys--Lleas 134
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 425 CAACGGAGCTTCACACAGAACTTCATCTGACGCCGCCAGCCTCACACAGACAGT 484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 134 ngllygly----- 136
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 485 TCGAAGATGCCAGTGTGCTTCGCTTCAGAGACTGGGCCAGACTGAGGGGTGGCAG 544
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 137 -----GluValAlaValAlphealaserArgAlaglyProGlyValCysTrpHispr 153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 545 AGGTCTTGTGCTTCATTAGCCCTAGCTCTGA----- 578
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153 osercysPheValCysPheThrCysasnGlnleuLeuValAspLeuIleTyrPheTyrG 173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 579 -----GAGAAATGCMAACTGACTTCAG 604
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 173 naspglyLysIleHiscysGlyArgHisHlsAlaGlnleuLysProArgCysSerAl 193
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 605 GATGTGAGAGACAAAGT----- 623
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193 acys-----AspGlnleuIlePheAlaAspGlnCysThrGlnAlaGlnGlyArgHisTr 211
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 624 -----CATTTCTGTGT-----TGGGAGACACTG-----CAGACTCCAC 658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 phlsmetlyshsPheCysCysleuGlnCys-GlnThrValleuGlyGlyArgGlyTyrI 231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 659 TGTGCCAGAGTTGAAGTCTTTTGTGTGCTCAAGTCTAGAGTCCCTTCTCGAATA 718
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 231 leuetylaspslylArgProPheCysGlyCysPheGlnserleuTyrAlaGlnTyrC 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 719 TACTGTGTGT 729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 251 ysglnthrcys 254
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

## RESULT 13



```

PRIORITY APPLICATION NUMBER: US 09/862,535
PRIORITY FILING DATE: 2001-05-21
PRIORITY APPLICATION NUMBER: US 60/205,961
PRIORITY FILING DATE: 2000-05-19
PRIORITY APPLICATION NUMBER: US 09/870,383
PRIORITY FILING DATE: 2001-05-29
PRIORITY APPLICATION NUMBER: US 60/207,506
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/860,821
PRIORITY FILING DATE: 2001-05-18
PRIORITY APPLICATION NUMBER: US 60/205,449
PRIORITY FILING DATE: 2000-05-19
PRIORITY APPLICATION NUMBER: US 09/870,110
PRIORITY FILING DATE: 2001-05-29
PRIORITY APPLICATION NUMBER: US 60/207,650
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/907,509
PRIORITY FILING DATE: 2001-07-16
PRIORITY APPLICATION NUMBER: US 60/218,385
PRIORITY FILING DATE: 2000-07-14
PRIORITY APPLICATION NUMBER: US 09/945,327
PRIORITY FILING DATE: 2001-08-31
PRIORITY APPLICATION NUMBER: US 60/229,425
PRIORITY FILING DATE: 2000-08-31
PRIORITY APPLICATION NUMBER: US 60/318,581
PRIORITY FILING DATE: 2001-09-10
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FASTSEQ Version 4.0
SEQ ID NO 17
LENGTH: 725
TYPE: PRT
ORGANISM: Homo sapiens
US-10-160-501-17

Alignment Scores:
Pred. No.: 27.1 Length: 725
Score: 78.50 Matches: 54
Percent Similarity: 40.40% Conservative: 47
Best Local Similarity: 21.60% Mismatches: 76
Query Match: 5.27% Indels: 73
DB: 9 Gaps: 15

US-09-763-902b-10 (1-830) x US-10-160-501-17 (1-725)
QY 41 CCGTCGAGCCCGTGGTCCCAAGCAGAGAAATACCTGTCGAGCCCTCCATCA 100
DB 391 ProserAsnAspPheMetProValAlaGlyProLeuValAlaHisLeuSerThrProArgAla 410
QY 101 GACCAGAGATGGCATCTGTGATTTCAAGACCTATGTGGATCAGGCTCGACAGCTGCT 160
DB 411 Asppro-----SerThrArgIleGluThrGlyValArgIleGly 423
QY 161 GAGAGATTGTCATGCTACTACACCAACAGATGATGAGCGCGCTTGTGTCGCCG 220
DB 424 AspGlu---ValSerValHisTyrAspPheMet-----IleAlaHis 436
QY 221 CTGTACATGGCAGCAGCCCGTGTGATGGCATGCTGTTTCAGACAGAGAAATCC 280
DB 437 Leu-----ValValTyr-----AlaAlaAspArgIleAlaAla 447
QY 281 TTGATGAGTTTGAATGTTGCTTCACGCGATTCACCAATACGCTGTGATGATGCTG 340
DB 448 LeuThrLys-----LeuArgTyrSerLeuArgIleTyrAsnIleVal 461
QY 341 CAGCCTGTTCATGATGAGCAGCAGCCAGCGATGCTGTTGTCATGCTGATGGA 400
DB 462 -----GlyLeuProThrAsnIleAspPheLeuLeuAsnLeuSerGly 475
QY 401 TCACTGAGATTGAG--GGGAAACAACAACGCGACTTC-----AACCAAGAC 445
DB 476 HisProGluPheGluAlaGlyAsnValHisThrAspPheIleProGluHisHisLysGln 495
QY 446 TTCATCCGACCGCCGCGAGCGCTCAGCCAGCAACAGACAGTGTGAAATGCGCA----- 496

```

```

DB 496 LeuLeuLeuSerArgLysAlaAlaAlaLysGluSerLeuCyGlnAlaAlaLeuGlyLeu 515
QY 497 -----AGTACCTGCTTC-----CGCTTCAG 517
DB 516 IleLeuLysGluLysAlaMetThrAspThrPheThrLeuGlnAlaHisAspIlePheSer 535
QY 518 GACTGGCCGACCTAGTGGCGGTGGCAGAGAGCTGCTGTTGCTGATTCAGCCCTACGCTGT 577
DB 536 ProPheSerSerSerSerGlyArgArgLeuAsnIleSerTyr----- 549
QY 578 AGAGAAATGCAACCTCGACCTCAAGATGTGAGAACACAGATTCATTTCTGTGTG 637
DB 550 -----ThrArgAsnMetThrLeuLysAspGlyLysAsnAsnValAlaIleAlaValThr 567
QY 638 CGGAGACACTGCAAGACTCCACTGTGCCGAGTTGA-----CTCTTTTGTGCTCA 691
DB 568 TyrAsnHisAspGlySerTyrSerMetGlnIleGluAspLysThrPhe-----Gln 584
QY 692 GTTCTAGGAGTCCCTTTCTGTAATATATAC 721
DB 585 ValLeuGly-----AsnLeuTyr 590

RESULT 15
US-10-146-473-50
Sequence 50, Application US/10146473
Publication No. US2003010888A1
GENERAL INFORMATION:
APPLICANT: Scanlan, Matthew
APPLICANT: Gout, Ivan
APPLICANT: Stockert, Elisabeth
APPLICANT: Gure, Ali
APPLICANT: Chen, Yao-Tseng
APPLICANT: Old, Lloyd
TITLE OF INVENTION: Breast Cancer Antigens
FILE REFERENCE: L00461/70130(JRV)
CURRENT APPLICATION NUMBER: US/10/146,473
PRIORITY FILING DATE: 2002-05-15
PRIORITY APPLICATION NUMBER: US 60/291,150
PRIORITY FILING DATE: 2001-05-15
NUMBER OF SEQ ID NOS: 82
SOFTWARE: PatentIn version 3.0
SEQ ID NO 50
LENGTH: 2971
TYPE: PRT
ORGANISM: Homo sapiens
US-10-146-473-50

Alignment Scores:
Pred. No.: 46.7 Length: 2971
Score: 78.50 Matches: 55
Percent Similarity: 32.17% Conservative: 19
Best Local Similarity: 23.91% Mismatches: 88
Query Match: 5.27% Indels: 68
DB: 9 Gaps: 7

US-09-763-902b-10 (1-830) x US-10-146-473-50 (1-2971)
QY 27 CTTCAACCTTACTTCCCTCCAGCCCTGTTCCCAAGCAGAGAAATACCTGTGTA 86
DB 1221 LeuAlaProValValProAlaIleAlaProGlyProProSerLeuGln-----ProSerGly 1238
QY 87 GCCCTCTTCATAGACACAGAGATGCAATGCTGTGATTTCAAGACCTATGATCAGG 146
DB 1239 Ala-----Ser 1240
QY 147 CTTGCAAGACTGCTGAGAGATTGTTCATATGCTTACTACCA----- 188
DB 1241 ProSerAlaSerAlaLeuThrLeuGlyLeuAlaThrAlaProSerLeuSerSerGln 1260
QY 189 -----CCATGATAGCGGC 203
DB 1261 ThrProGlyHisProLeuLeuLeuAlaProThrSerSerHisValProGlyLeuAsnSer 1280

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 19:26:18 ; Search time 27.5127 Seconds  
(without alignments)  
5800.355 Million cell updates/sec

Title: US-09-763-902b-10  
Perfect score: 1489  
Sequence: 1 tgaacttgacatcagctg.....ataatttaataacacatg 830

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p.model -DEV=xip  
-O=/cgn2.1/USPTO.spool/US09763902/runal.24062003.130341.12467/app\_query.fasta.1.2254  
-DB=PIR\_73 -QWMT=fastan -SUFMT=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0 -LIST=45  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USFR=US09763902.cgn.1.1.108.grunt.24062003.130341.12467 -NCPU=6 -ICPU=3  
-NO\_MMP -IARECOVERY -NEG\_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	8.3	127	2	S00751 placental protein
2	118	7.9	522	2	H86248 protein T23J18.22
3	117.5	7.9	126	2	B86405 probable nuclear t
4	116.5	7.8	125	2	S50467 nuclear transport
5	107	7.2	123	2	T38039 probable nuclear t
6	106	7.1	537	2	T23479 hypothetical prote
7	105	7.1	122	2	H86398 protein F17L21.10
8	103.5	7.0	133	2	T23921 hypothetical prote
9	91.5	6.1	434	2	T40805 probable RNA-bindi
10	90.5	6.1	685	1	EFBY52 suppressor 2 prote
11	88.5	5.9	353	2	B72468 hypothetical prote
12	87	5.8	485	2	T49237 hypothetical prote
13	86.5	5.8	1309	2	T00078 probable RNA-dirc
14	86	5.8	328	2	H75251 glycerol-3-phospha

C	15	86	5.8	7962	2	I38346
C	16	85.5	5.7	915	2	S54485
C	17	83.5	5.6	197	2	A53126
C	18	83.5	5.6	355	2	JN0795
C	19	83.5	5.6	475	1	JN0327
C	20	83.5	5.6	475	2	E91079
C	21	83.5	5.6	475	2	F85924
C	22	83	5.6	427	2	E96716
C	23	82.5	5.5	458	2	T27007
C	24	81.5	5.5	458	2	A40392
C	25	81.5	5.5	573	2	A33533
C	26	81	5.4	309	2	T25800
C	27	81	5.4	1198	2	T28678
C	28	80.5	5.4	152	2	C72653
C	29	80.5	5.4	319	2	C72729
C	30	80.5	5.4	423	2	A48152
C	31	80.5	5.4	428	2	F86270
C	32	80.5	5.4	773	2	A82998
C	33	80	5.4	143	2	E72699
C	34	80	5.4	555	1	RGASMA
C	35	79.5	5.3	135	2	G84212
C	36	79.5	5.3	185	2	T15623
C	37	79	5.3	515	2	S63382
C	38	78.5	5.3	364	2	D82424
C	39	78.5	5.3	455	2	B82525
C	40	78.5	5.3	546	2	F70866
C	41	78.5	5.3	622	2	T24632
C	42	78.5	5.3	649	1	H64476
C	43	78	5.2	296	1	UC7268
C	44	78	5.2	316	1	S27981
C	45	78	5.2	469	2	A57531

## ALIGNMENTS

elastic titin - hu  
CES1 protein - yea  
syndecan-4 precurs  
butyrate kinase (E  
sulfate adenylyl tr  
hypothetical prote  
hypothetical prote  
probable RNA-bind  
alpha-2-adrenergic  
cell surface glyco  
C2H2-type zinc fin  
polyketide synthas  
hypothetical prote  
hypothetical prote  
zinc finger protei  
hypothetical prote  
probable oxidoredu  
hypothetical prote  
regulatory protein  
hypothetical prote  
hypothetical prote  
luxp protein VCA07  
conserved hypothet  
probable maltase -  
hypothetical prote  
endopeptidase La h  
paired-type homeod  
homoserine kinase  
EGR alpha transcri

RESULT 1	S00751	placental protein 15 - human
N/Alternate names:	Pp15	
C/Species:	Homo sapiens (man)	
C/Date:	31-Dec-1988	#sequence-revision 31-Dec-1988 #text-change 24-Sep-1999
C/Accession:	S00751	
R/Grundmann, U.;	Neirlich, C.;	Rein, T.; Lottspeich, F.; Kuepper, H.A.
Nucleic Acids Res.	16	4721, 1988
A>Title:	Isolation of cDNA coding for the placental protein 15 (Pp15).	
A/Reference number:	S00751; MUID:88247772; PMID:3380696	
A/Accession:	S00751	
A/Molecule type:	mRNA	
A/Cross-references:	EMBL:X07315; NID:g35578; PIDN:CAA30278.1; PID:g35579	
C/Superfamily:	Yeast nuclear transport factor NTF2	
Alignment Scores:		
Pred. No.:	0.000455	Length: 127
Score:	124.00	Matches: 31
Percent Similarity:	43.59%	Conservative: 20
Best Local Similarity:	26.50%	Mismatches: 54
Query Match:	8.3%	Indels: 12
DB:	2	Gaps: 3
US-09-763-902b-10 (1-830) x S00751 (1-127)		
QY	167	TTTGTCAATGCTGTACTACACCAATGATGACGGCGGCTGTCTGTCGCCCTGTAC 226
DB	14	PheileglnhstyttyglInleupheapsnspahynglnleuglYAlaIlelyr 33
QY	227	ATGGCACACCACTGCTGTGGAATGCAATGCTGTTTCAGACGAAGAATCTTGAGT 286
DB	34	IleaspalasercysleuthrtrpbluglnglnphleglnclYlyeAlaIlel 53
QY	287	GAGTTTGTGAATGTGCTTCAGCAGATTCACAAATCAGCGTGTA-----GACTGC 340
DB	54	GtulyseuserSerleupProphlelnlyslleglnhstserllethrAlaIlelnspHls 73

OY 443 ATCTCG 454  
 111  
 Db 438 H1SLeu 439  
 111  
 111

RESULT 3  
 B86405  
 Probable nuclear transport factor 2 (imported) - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: B86405  
 R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Aloni  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
 Nauen, N.F.; Hughes, B.; Huizart, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maitl, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Sano, T.; Sano, T.; Sano, T.; Sano, T.;  
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talid  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A6141; MUID:21016719; PMID:11130712  
 A:Accession: B86405  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-126 <STO>  
 A:Cross-references: GB:AE005172; NID:g11024877; PIDN:AG26961.1; GSPDB:GND0141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:  
 Pred. No.: 0.00207 Length: 126  
 Score: 117.50 Matches: 37  
 Percent Similarity: 47.97% Conservative: 22  
 Best Local Similarity: 30.08% Mismatches: 49  
 Query Match: 7.89% Indels: 15  
 2 Gaps: 5

US-09-763-902B-10 (1-830) x B86405 (1-126)  
 OY 158 GCTGAGGAGTTGTTCATGTTCTACTACACCCATGATAGCGCGCGTTGCTGTC 217  
 Db 10 SerLysAlaIheValGluHISLTYTTrSerThrPheAspArgValGlyLeuAla 29  
 OY 218 CGCCTGATCTGGGACACACCCCTGCTGATGAGTGGATGCTGTTTCAAGGACAA 277  
 Db 30 GlyLeuTyrGlnGluAlaSerMetLeuThrPheGluGlnGlnLysIleGlnGlyAlaGln 49  
 OY 278 TCCCTAGAGGAGTTTGTGAATGTTCCGT-----TCCAGCGAGTTCCCAATCAGCGCTG 331  
 Db 50 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnLysIleHisIleSerThr 69  
 OY 332 GTAGACTGCCAGCGCTTTCATGATGAAGCACACCAAGCCAGCAGCGCTGCTGTTTC 391  
 Db 70 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 84  
 OY 392 ATCTGTGGATCAGTGAAGTTTGAGGGGAACAAACAAGG---GACTTCAACACAGACGTC 448  
 Db 85 ValSerGlyAsnLeuGlnLeuAlaGlyGlnGluHISAlaLeuLysPheSerGlnMetPhe 104  
 OY 449 ATCTGT-----ACCGCCAGCGCTTACCCAGACACAGTGTGGAAGATGCAAGTGAC 502  
 Db 105 HisLeuMetProThrProGlnGlySer-----PheTyrValPheAsnAsp 119  
 OY 503 TGCTTCCGC 511  
 Db 120 IlePheArg 122

RESULT 4  
 S50467  
 nuclear transport factor NTF2 - yeast (Saccharomyces cerevisiae)  
 ;Alternate names: protein YFR004

C/Species: *Saccharomyces cerevisiae*

C>Date: 28-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 24-Sep-1999  
 C/Accession: S50467; S41793; S72237  
 R: Dietrich, F.S.  
 Submitted to the EMBL Data Library, December 1994  
 A/Description: The sequence of *S. cerevisiae* cosmid 9537, 9581, 9495, 9867, and lambda  
 A/Reference number: S50459  
 A/Accession: S50467  
 A/Molecule type: DNA  
 A/Residues: 1-125 <DIE>  
 A/Cross-references: EMBL:U18778; NID:603592; PIDN:AAB64542.1; PID:6603601; MIPS:YER009W  
 R: Haarer, B.K.; Petzold, A.S.; Brown, S.S.  
 Submitted to the EMBL Data Library, July 1993  
 A/Description: Identification of mutations that are synthetically lethal with altered ye  
 A/Reference number: S41793  
 A/Accession: S41793  
 A/Molecule type: DNA  
 A/Residues: 1-84 <HAA>  
 A/Cross-references: EMBL:L22204; NID:9347714; PIDN:AAB49379.1; PID:9347715  
 R: Haarer, B.K.; Corbett, A.; Kweon, Y.; Petzold, A.S.; Silver, P.; Brown, S.S.  
 Genetics 144, 495-510, 1996  
 A/Title: SEC3 mutations are synthetically lethal with profilin mutations and cause defec  
 A/Reference number: S72237; MUID:97044444; PMID:8889515  
 A/Accession: S72237  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-84 <HAW>  
 A/Cross-references: EMBL:L22204; NID:9347714; PIDN:AAB49379.1; PID:9347715  
 C/Genetics:  
 A/Gene: SGD:NTE2  
 A/Cross-references: SGD:S000811; MIPS:YER009W  
 A/Map position: 5R  
 C/Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:  
 Pred. No.: 0.00261 Length: 125  
 Score: 116.50 Matches: 42  
 Percent Similarity: 44.85% Conservative: 19  
 Best Local Similarity: 30.88% Mismatches: 54  
 Query Match: 7.82% Indels: 21  
 Gaps: 7

US-09-763-902b-10 (1-830) x S50467 (1-125):

```

QY 116 TCTGTGATTTCAAGCTATGATGATCAGCGCTGAGAGTTCAT 175
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 SerLeuAspPheAsnThr-----LeuAlaGlnAsnPheThrGln 14
QY 176 GTCTACTACACCAATGATTAAGCGCGCGTTCCTGCTCCGCTGACATGGCACA 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 15 PheTYTAsnGlnPheAspThrAspArgSerGlnLeuGlnLeuTYrArgAsnGlu 34
QY 236 GCCACCCGTGTCGGAATGCAATGCTGTTTCAGACAGAACATCCTGAGTGTTC 295
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 35 SerMetLeuThrPheGlnThrSerGlnLeuGlnGlnAlaLysAspIleValGlnLysLeu 54
QY 296 GAAATGTCCTCCAGCGAGTCCAA-----ATGACGCTGGATGATCCAGCGCTGT 349
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 55 ValSerLeuProPheGlnLysValGlnHisArgIleThrThrLeuAspAlaGlnPro- 73
QY 350 CATGATGAGCCACACCAAGCAGACACGAGCTGTTGTCATCTGTGATCA----- 403
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 -----AlaSerProAsn---GlyAspValLeuValMetIleThrGlyAspLeuLeu 89
QY 404 GTGAAGTTGAGGGAGCAACAACAAGGACTTCACCAAGTTCATCTGACCGCCAG 463
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 90 IleAspGlnGlnLysAsnProGlnArg---PheSerGlnValPheHisLeuIle----- 106
QY 464 GCCTACCCGACACACAGTGTGAAGATCGCAAGTACTGCTTCGCC 511
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 107 -----ProAspGlyAsnSerTYrTYrValPheAsnAspIlePheArg 120

```

probable nuclear transport factor 2 - fission yeast (*Schizosaccharomyces pombe*)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 15-Sep-2000  
 C/Accession: T38039; T37728  
 R: McDougal, R.C.; Rajandream, M.A.; Bartell, B.G.; Whitehead, S.; Churcher, C.M.  
 Submitted to the EMBL Data Library, August 1999  
 A/Reference number: T21764  
 A/Accession: T38039  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-123 <MCDO>  
 A/Cross-references: EMBL:AL109951; PIDN:CAB53052.1; GSPDB:GN00066; SPDB:SPAC1B9.01C  
 A/Experimental source: strain 972h-1; cosmid c1B9  
 R: Devlin, K.; Churcher, C.M.; Bartell, B.G.; Rajandream, M.A.; Walsh, S.V.  
 Submitted to the EMBL Data Library, December 1995  
 A/Reference number: T21740  
 A/Accession: T37728  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 57, 77-123 <DEV>  
 A/Cross-references: EMBL:Z68197; PIDN:CAA92380.1; GSPDB:GN00066; SPDB:SPAC1B9.03C  
 C/Genetics:  
 A/Gene: SPDB:SPAC1B9.01C  
 A/Map position: 1  
 A/Intons: 3/1; 30/3; 54/3; 88/3  
 C/Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:  
 Pred. No.: 0.0238 Length: 123  
 Score: 107.00 Matches: 35  
 Percent Similarity: 46.28% Conservative: 21  
 Best Local Similarity: 28.93% Mismatches: 53  
 Query Match: 7.19% Indels: 12  
 Gaps: 4

US-09-763-902b-10 (1-830) x T38039 (1-123):

```

QY 158 GCTGAGAGTGTGTCATGTCATACACCAATGATTAAGCGCGCTTCCTGCTCC 217
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 8 AlaThrGlnPheThrGlnPheTYrTYrGlnThrPheAspSerAspArgSerGlnLeuSer 27
QY 218 CGCCTGATACAGGACACGACCGCCTGCTGATGAATGCAATGCTGTCAGGACAAGAA 277
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 28 SerLeuTYrArgGlnGlnLeuSerMetLeuSerPheGlnGlnAlaGlnLeuGlnGlnTYrThrLys 47
QY 278 TCCCTGAGTGTGTTTGAATGTTGCCCTTCACGAGATTCCA-----ATCAGCGTG 331
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 48 AlaIleValGlnLysLeuValSerLeuProPheGlnArgValGlnHisArgIleSerThr 67
QY 332 GTAGACTGCCAGCCTGTCATGATGAACCAACCAAGCCAGACACGAGTCTGTTGTC 391
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 68 LeuAspIleGlnPro-----ThrGlyThrThrGlySerValIleValMet 82
QY 392 ATCTGTGATCATGTAAGTT-----GAGGGAACAACAAGGAGCTTCACCAAGTTC 448
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 83 ValTYrGlnGlnLeuLeuLeuAspGlnGlnMetAlaGlnArgTYrSerGlnValPhe 102
QY 449 ATCTGACCGCCACGAGCTCAGACCAACAGTGTGAAGATCGCAAGTACTGCTTC 508
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 103 HisLeuVal-----AsnAsnAsnGlyAsnTYrTYrValLeuAsnAspLeuPhe 118
QY 509 CGC 511
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 119 Arg 119

```

RESULT 6  
 T23479  
 hypothetical protein K08F4.2 - *Caenorhabditis elegans*  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C/Accession: T23479  
 R: Hamby, C.  
 Submitted to the EMBL Data Library, January 1996







[illegible]

613 CCTGACATCCCTTGAGAGCTGCAGGTTTGATTTCTGTACAGAGCTAGGCGCTGAATGACCA 554  
 148 ProProLeuLeuGluProSerGlyIleAsnTyrSerValTyrTyrGluLeuLeuGluIle 167  
 553 AAGAGACCT-----CTGCCACCCCACTACTACTGGCCCAAGTCCCTGGAAACCGAA----- 506  
 168 GluProProGluTyrPleuProAspSerLeuAlaSerThrCysMetGlnCysSerThrPro 187  
 505 -----GCAGTCACTTCGATCTTCCACACAGCT---GTTGCTGGCTAGAGCGCTGGCGGCT 455  
 188 PheThrAlaIleThrCysGlyAlaGlnHisCysAlaPheCysGly-----Gly-I1 204  
 454 CAGGATGAAGTCTGTGTTGAAGTCCGCTGTTGTTTCCCTCAACTTC----- 406  
 204 ePheCysArgAsnCysSerTyrGlyAlaGlyCysLeuMetProSerAlaPheArgGluArgAs 224  
 405 -----

Db 224 nProglArgValCysaspSerCysTyrGluArgLeuAspProLeuIncCysValLeu1 244  
 QY 396 -----CAGATGACACAGACCGGTGTC---TGGCTGGTGT 362  
 Db 244 eaasSerIleSerAsnAlaValGlnValAlaLysHisAspValValAspTrp----- 261  
 QY 361 GGCTTCATCATGACAGCGTGGCAGCTGATCCAGCTGATTGGACCTGCTGAGAGCAA 302  
 Db 262 ----ThCysSerArgGlyTrpLeuAsnLeuPro---ValGlyLeuSerMetGluAspG1 279  
 QY 301 CATTTCAAAA-----AACTCACTCAAGAGATTCTGT-----CCTGA 266  
 Db 279 uileTyrLysAlaAlaAsnThrLeuArgGlyTyrCysGlnValAlaArgLeuAspProG1 299  
 QY 265 AACAGCATTCGCA 253  
 Db 299 uLysSerIlePro 303

## RESULT 13

T00078  
 Probable RNA-directed DNA polymerase (EC 2.7.7.49) - Chlorella vulgaris retrotransposon  
 N:Alternate names: reverse transcriptase  
 C:Species: Chlorella vulgaris  
 C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
 C:Accession: T00078  
 R:HisashiYama, T.; Noutoshi, Y.; Fujie, M.; Yamada, T.  
 EMBL J. 16, 3715-3723, 1997  
 A:Title: zepp, a LINE-like retrotransposon accumulated in the Chlorella telomeric region  
 A:Reference number: 214092; MUID:97361851; PMID:9218812  
 A:Accession: T00078  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1309 <HIG>  
 A:Cross-references: EMBL:AB008896; NID:g3063386; PIDN:BAA25763.1; PID:g3073779  
 A:Experimental source: strain C-169  
 C:Genetics:  
 A:Map position: 5  
 A:Mobile element: retrotransposon zepp  
 C:Keywords: nucleotidyltransferase

## Alignment Scores:

Pred. No.: 3 93 Length: 1309  
 Score: 86.50 Matches: 71  
 Percent Similarity: 29.35% Conservative: 20  
 Best Local Similarity: 22.90% Mismatches: 99  
 Query Match: 5.81% Indels: 120  
 Gaps: 16

US-09-763-902b-10 (1-830) x T00078 (1-1309)

QY 656 GGAGCTCTGCACTGTCTCCGACAACAGAA-----ATGAACTGTGTTCCTCAAT 606  
 Db 968 GlyThrCysGlnArgGlyGlySerThrSerGlySerTrpGluAspSerCysSerSerPro 987  
 QY 605 CCTTGAGAGTCGAGGATTGCAATTCTCTACAGAGCTAGGCTGAA----- 561  
 Db 988 ProSerGlnArgSer-----MetTrpArgProArgSerAlaTyrGlySerThrCys 1004  
 QY 560 -----TGAACCAAGAGACCTGTGCCCCCCCC 534  
 Db 1005 MetProThrSerSerCysGlyProIleThrCysGlySerArgGlySerAlaThrThr 1024  
 QY 533 ACTAGCTGGCCAGT-----CCTGGAAGCGGAACCATCTCCATCTCCACACTG 480  
 Db 1025 GlnSerTrpValSerHisCysSerSerThrGlySerSerArgMetGlnSerAlaAlaLeu 1044  
 QY 479 TGT-----TGCTGG 471  
 Db 1045 CysLeuGlyGlyLeuGlyGlnAlaTrpProGlyGlnAspGlnProProAlaArgCysArg 1064  
 QY 470 GTAGAGCTGGCGGCTCAGATGAAGTCTGGTTAGTCCCGTTGTTTCCCTCAA 411

Db 1065 AlaArgProGlyAlaSerGly---SerAspGly----- 1074  
 QY 410 ACTTCACCTGATCCACAGATGACACAGACCGGTGCTGCTGCTGCTGCTTCATCAT 351  
 Db 1075 -----AlaArgAlaArgAlaHisPro---SerGlyCysLeuAlaSerSerLeu 1089  
 QY 350 GAACAGCGTGGCAGCTGTACACACCGCTGATTGGAACTGCTGGAAGACACATTTCAAA 291  
 Db 1090 GluSerAlaAsn-----ProGlyAlaAlaThrGlyArgTyrValValArgIleSerAsp 1107  
 QY 290 ACTTACCTCAAGATTTCTTTCCTGAAACAG-----CAT 258  
 Db 1108 ProSerAspArgValCysLeuAlaGlnProLysValAlaProThrSerGlnProHis 1127  
 QY 257 TGCCATTCACAGACAGGAGTGGCTGTGC-----CCATTCACAGGC 219  
 Db 1128 SerLysGlnHisArgSerLeuLeuCysGlnTrpThrLysAlaLeuGlnProCysThrHis 1147  
 QY 218 GGGACAGCAACAGCGCGCGCTTATCCATGG----- 189  
 Db 1148 SerPheLeuArgProArgSerLeuHisThrPheSerThrGlyThrSerProGlyProGly 1167  
 QY 189 ----- 189

Db 1168 IleHisAlaAlaHisValSerSerAlaSerThrGlnSerSerCysThrGlnSerSer 1187  
 QY 188 -----TGGTGTAGTACATTCGACAACTCTCAG----- 159  
 Db 1188 ThrCysArgTrpSerSerArgGlnGlySerSerGlnGlySerSerTrpGlySer 1207  
 QY 158 CAGCTGCGAGCGCTGATCCACATAGAGCTTGAATGCCACAGATCCATCTGCTGCTTA 99  
 Db 1208 SerSerCysArgSerHisCysHisHisArg---ArgProHisLeuTyrPheLeuCysLeu 1226  
 QY 98 TGAAGAGAGCGCTCCACAGGGGTATTCCCTGCTGCTG-----GGAACCGAGCGCTGC 45  
 Db 1227 Trp-----GlySerGlnValGlnAspProGlnAlaTrpGlyLeuGlyThrArgProAla 1244  
 QY 44 AGGAGATTAAGTTGACAGAGCCAGC 17  
 Db 1245 HisGluPheValVal-ArgGlnAlaThr 1253

## RESULT 14

H75251  
 glycerol-3-phosphate dehydrogenase, MND(+)-dependent - Deinococcus radiodurans (stra  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: H75251  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.  
 M.; Shen, M.; Yamathavan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: H75251  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-328 <WHI>  
 A:Cross-references: GB:AE002091; GB:AE000513; NID:g6460446; PIDN:AAF12158.1; PID:g64  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR2621  
 A:Map position: 1  
 C:Superfamily: glycerol-3-phosphate dehydrogenase

## Alignment Scores:

Pred. No.: 3 64 Length: 328  
 Score: 86.00 Matches: 62  
 Percent Similarity: 39.65% Conservative: 28  
 Best Local Similarity: 27.31% Mismatches: 77  
 Query Match: 5.78% Indels: 60  
 Gaps: 11



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:24:08 ; Search time 12.6012 Seconds  
(without alignments)  
5463.815 Million cell updates/sec

Title: US-09-763-902b-10  
Perfect score: 1489  
Sequence: 1 tgaatttgatcattcacgty.....ataattaataacacatg 830

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 112892 segs, 41476328 residues  
Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame.n2p.model -DEV=rlp  
-Q=/cg92\_1/USPTO.spool/US09763902/runat\_24062003\_130340\_12418/app\_query.fasta.1.2254  
-DB=SwissProt.40 -QEXT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTMT=ptc -NCBM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US09763902.ecgn.1.1.46 -tnameat\_24062003\_130340\_12418 -KCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	732	49.2	140	1	NXT1_HUMAN
2	726	48.8	140	1	NXT1_MOUSE
3	584	39.2	142	1	NXT2_HUMAN
4	308	20.7	133	1	NXT1_DROME
5	206.5	13.9	137	1	NXT1_CAEEL
6	124	8.3	127	1	NXT2_HUMAN
7	123	8.3	127	1	NXT2_XENLA
8	118.5	8.0	126	1	NXT2_CANAL
9	117.5	7.9	125	1	NXT2_ARATH
10	116.5	7.8	125	1	NXT2_YEAST
11	115.5	7.8	125	1	NXT2_HUMAN
12	115.5	7.8	122	1	NXT2_MOUSE
13	107	7.2	122	1	NXT2_ORYSA
14	107	7.2	122	1	NXT2_SCHPO
15	105.5	7.1	123	1	NXT2_HUMAN
16	103.5	7.0	123	1	NXT2_CAEEL
17	102.5	6.9	124	1	NXT2_NEUCR
18	101.5	6.8	124	1	NXT2_MOUSE

19	91.5	6.1	434	1	G3BP_SCHPO	094260 schizosacch
20	90.5	6.1	685	1	ERF2_YEAST	P05453 saccharomyc
21	90	6.0	479	1	KLFA_MOUSE	O89091 mus musculu
22	87	5.8	351	1	ROM1_BOVIN	P52205 bos taurus
23	86	5.8	328	1	GPDA_DEIRA	O91716 delinococcus
24	85.5	5.7	915	1	ZDS1_YEAST	P50111 saccharomyc
25	83.5	5.6	197	1	SDC4_CHICK	P49416 gallus gall
26	83.5	5.6	355	1	BURK_CLOVE	O05619 elostriidum
27	83.5	5.6	475	1	CYSN_ECOLI	P23845 escherichia
28	81.5	5.5	573	1	C114_MOUSE	P19467 mus musculu
29	80.5	5.4	423	1	GFI1_RAT	O07120 rattus norv
30	80	5.4	555	1	WETA_EBNTI	P22022 emeticella
31	79.5	5.3	515	1	DABI_RHIL0	P58210 rhizobium
32	79	5.3	515	1	YN8T_YEAST	P53741 saccharomyc
33	79	5.3	1042	1	CORI_HUMAN	O9Y5G5 homo sapien
34	78.5	5.3	364	1	LUXP_VIBCH	O9K116 vibrio chol
35	78.5	5.3	365	1	FCGN_MOUSE	O61559 mus musculu
36	78.5	5.3	622	1	YRT1_CAEEL	O10044 caenorhabdi
37	78.5	5.3	649	1	LONH_MERJA	O58812 methanococc
38	78.5	5.3	725	1	MCCA_HUMAN	O96193 homo sapien
39	78	5.2	480	1	KLFA_HUMAN	O13118 homo sapien
40	78	5.2	1098	1	TSU1_YEAST	P38437 saccharomyc
41	77.5	5.2	765	1	SIM1_MOUSE	O61045 mus musculu
42	77.5	5.2	2738	1	PGCV_RAT	O98147 rattus norv
43	77	5.2	838	1	VGLH_HSV11	P06477 herpes simp
44	76.5	5.1	353	1	CUPS_GALME	O24998 galliera me
45	76.5	5.1	458	1	A2AC_MOUSE	O01337 mus musculu

## ALIGNMENTS

## RESULT 1

ID NXT1\_HUMAN STANDARD: PRT: 140 AA.

AC Q9UKK6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE NTF2-related export protein 1 (p15 protein).

CN NXT1.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A. AND FUNCTION.

RX MEDLINE=20036817; PubMed=10567585;  
RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;

RT "Identification of an NTF2-related factor that binds Ran-GTP and  
RT regulates nuclear protein export."

RL Mol. Cell. Biol. 19:8616-8624(1999).

RN [2]  
RP SEQUENCE FROM N.A.

RX MEDLINE=21638749; PubMed=11780053;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggaley C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,

RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Butler A.P., Carter C., Carter N.P.,

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Cooley V.E., Collier R.E., Connor R.E., Corby N.R.,

RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

RA Griffiths D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond E., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,

RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,

RA Levasseur M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConachie L.J., McKay K., McKerrry A.A.,

RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
RA Rice C.M., Ross M.T., Scott C.E., Sehara H.K., Shownkeen R., Sims S.,

RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Treacy A., Toman A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.,  
 RT "The DNA sequence and comparative analysis of human chromosome 20.",  
 RL Nature 414:865-871(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung, Placenta, and Uterus;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP FUNCTION.  
 RX MEDLINE=20307861; PubMed=10848583;  
 RA Ossareh-Nazari B., Watson C., Black B.E., Levesque L., Paschal B.M.,  
 RA Dargemont C.;  
 RT "Rangit-binding protein Nxtl facilitates nuclear export of different  
 RT classes of RNA in vitro.";  
 RL Mol. Cell. Biol. 20:4562-4571(2000).  
 RN [5]  
 RP FUNCTION.  
 RX MEDLINE=21160285; PubMed=11259602;  
 RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
 RA Paschal B.M., Rekosh D., Hammarskjold M.L.;  
 RT "Nxtl (p15) is a crucial cellular cofactor in TAP-dependent export of  
 RT intron-containing RNA in mammalian cells.";  
 RL Mol. Cell. Biol. 21:2545-2554(2001).  
 CC CC  
 CC -1- FUNCTION: Stimulator of protein export for NES-containing  
 CC proteins. Also plays a role in the nuclear export of U1 snRNA,  
 CC tRNA, and mRNA.  
 CC -1- SUBUNIT: Preferentially binds Ran-GTP. Associates with Nxf1, Nxf2,  
 CC Nxf3 and Nxf5.  
 CC -1- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and  
 CC the cytoplasm.  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC CC  
 DR EMBL: AF156957; AAD54942.1;  
 DR EMBL: AL096677; CAC03437.2;  
 DR EMBL: BC000759; AAH00759.1;  
 DR EMBL: BC002687; AAH02687.1;  
 DR EMBL: BC003029; AAH03029.1;  
 DR EMBL: BC003410; AAH03410.1;  
 DR Gene: HGNC:15913; NMTL.  
 DR MIM: 605811;  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
 KW Transport; Protein transport; Nuclear protein.  
 FT DOMAIN 16 135 NTF2  
 SEQUENCE 140 AA: 15847 MW: 356FA86AC3944594 CRC64:

Alignment Scores:  
 Pred. No.: 4,87e-65 Length: 140  
 Score: 732.00 Matches: 140  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 49.16% Indels: 0  
 DB: 1 Gaps: 0

US-09-763-902b-10 (1-830) x NMTL\_HUMAN (1-140)

OY 110 ATGCGATGTGGATTTCAAGACCATGTGATCAGCGCTGCAGAGCTGTGAGAGATT 169  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 1 MetalaserValasphelelysthrValaspinalacyArgAlaIacInclue 20  
 OY 170 GTCAATGCTCTACACCATGATGATGAGGGGGCTTGCTGCTCCGCTGACATG 229  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 21 ValasnValytrThrThmetaspysrghArgLeuLeuSerArgLeuYrMet 40  
 OY 230 GGCACAGCCACCCGTGCTGGAATGCGCAATCTGTTTCAGACAAAGATCCCTGATG 289  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 41 GlyThrAlatrThrLeuValTrpAsnGlyAsnAlaValSerGlyIngluSerleu 60  
 OY 290 TTTTGAATGTGCTTCAGGAGTTCACCAATCAGCTGAGAGTCCACCTGT 349  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 61 PhepHeuMetleuProserSerGluPheGlnIleSerValValaspCysGlnProval 80  
 OY 350 CATGATGAGCCACACCAAGCCACACCGGCTCTGCTGCTGCTGCTGCTGCTGAG 409  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 81 HisaspGluAlatrProserGlnThrThrValleuValIleCysGlySerVallys 100  
 OY 410 TTGAGGGGGAACAACAAGGAGCTTCACCAAGCTTCAGCTGAGCCGCTCA 469  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 101 PhegluGlyAsnLysGlnArgAspPheAsnGlnAsnPheIleuThrAlaGlnAlaser 120  
 OY 470 CCCAGCACACACAGTGTGGAATGCGCAATGCTGCTGCTGCTGCTGCTGCTGAG 529  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 121 ProserAsnThrValTrpLysIleAlaserAspCysPheArgPheGlnAspTrpAlaser 140  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2  
 NMTL\_MOUSE STANDARD: PRT: 140 AA.  
 ID NMTL\_MOUSE  
 AC 0902V9; G9DIP7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE NTF2-related export protein 1.  
 GN NMTL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20036817; PubMed=10567585;  
 RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;  
 RT "Identification of an NTF2-related factor that binds Ran-GTP and  
 RT regulates nuclear protein export.";  
 RL Mol. Cell. Biol. 19:8616-8624(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aikawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Satou T., Okazaki Y., Gojopori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nakadoi T., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,  
 RA Blake J., Botfield D., Boujanga N., Carninci P., de Bonaldo M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereis P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz G., Whitaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC CC  
 CC -1- FUNCTION: Stimulator of protein export for NES-containing  
 CC proteins. Also plays a role in the nuclear export of U1 snRNA,  
 CC tRNA, and mRNA (By similarity).

CC -1- SUBUNIT: Preferentially binds Ran-GTP. Associates with NXF1, NXF2  
 CC and NXF3.  
 CC -1- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and  
 CC the cytoplasm.  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF156958; AAD54943.2;  
 DR EMBL: AK003254; BAB2670.1;  
 DR MGD: MG1:1929619; Nxtl.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2; 1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
 KW Transport; Protein transport; mRNA transport; Nuclear protein.  
 FT DOMAIN 16  
 FT CONFID 100 135 NTF2.  
 FT K -> N (IN REF. 2).  
 SQ SEQUENCE 140 AA; 15847 MW; 358EAFDC19FE4594 CRC64;

Alignment Scores:  
 Pred. No.: 1,93e-64 Length: 140  
 Score: 726.00 Matches: 138  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 98.57% Mismatches: 0  
 Query Match: 48,76% Indels: 0  
 DB: 1 Gaps: 0

US-09-763-902b-10 (1-830) x NMT1\_MOUSE (1-140)

QY 110 ATGCATCTGTGATTTCAAGACCTATGTGATCAGACCCGACAGAGCTGTGAGAGATT 169  
 |||||  
 Db 1 MetAlaSerValAspPheIytrThrValAspGlnAlaSerAlaAlaGluGluPhe 20  
 QY 170 GTCAATGTCTACTACACACCATGATAGCGGCGGGTGTGTGTCGCCCTGTACATG 229  
 |||||  
 Db 21 ValAsnValAltyrTrhTrhMetAspIyAlrGArgrgLeuLeuSerArgrLeuTyMet 40  
 QY 230 GGCACAGCCACCTCGTGTGATGCGATGCTGTTTCAGACAGCAAGATCTTGAGTGA 289  
 |||||  
 Db 41 GlyThralathrLeuValtrPaangIyAsnAlaValSerIyngIngluSerLeuSerGlu 60  
 QY 290 TTTTGAATGTGCTTCACGAGTTCACATGAGCTGAGTGTGAGTGTGAGTGTGAGT 349  
 |||||  
 Db 61 PhePheGluMetLeuProSerSerGluPheGlnIleSerValValAspCysGlnProVal 80  
 QY 350 CATGATGAAGCCACACACACACACGAGTGTGCTGTGTGTGATCTGTGATCAGTGA 409  
 |||||  
 Db 81 HisAspAspAlathrProSerGlnTrhThrValLeuValValIleCysGlyTrhValLys 100  
 QY 410 TTTCAGGAGGACAAACACGCGACTTCACACAGACTTCATCTGACGCCAGCCCTCA 469  
 |||||  
 Db 101 PheGlnGlyIyAsnIySglArgrAspPheAsnGlnAsnPheIleLeuThralaGlnAlaSer 120  
 QY 470 CCAAGACACACACTGTGGAAGATCCGACAGTGTGCTTCGCTTCAGAGTGTGAGTGA 529  
 |||||  
 Db 121 ProSerAsnThrValTrpIyIleAlaSerAspCysPheArgrPheGlnAspTrpAlaSer 140

RESULT 3  
 NMT2\_HUMAN  
 ID NMT2\_HUMAN STANDARD; PRT; 142 AA.  
 AC Q9NPB8; Q9NPB8; Q9NPB8; Q9NPB8; Q9NPB8; Q9NPB8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE NTF2-related export protein 2 (p15-2 protein) (DC9) (BM025).  
 GN NMT2.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE-Testis;  
 RX MEDLINE=20528640; Pubmed=11073998;  
 RA Herold A., Suyama M., Rodrigues J.P., Braun I.C., Kutay U.,  
 RA Carmo-Fonseca C., Bork P., Izauralde E.,  
 RT "MAP (NXF1) belongs to a multigene family of putative RNA export  
 RT factors with a conserved modular architecture".  
 RL Mol. Cell. Biol. 23:8996-9008(2000).  
 RN (2)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Kang Y., Cullen B.,  
 RT "p15-2, a homologous protein of p15, interacts with Tap".  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Li Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.,  
 RT "Novel genes expressed in human dendritic cell".  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE-Ovarian carcinoma;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Nagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakami K., Iishi S., Kawai Y.,  
 RA Wakamatsu J., Wakamatsu A., Nakamura Y., Nagahara K., Masuo Y.,  
 RA Minomiya K., Iwayanagi T.,  
 RT "NEDD human cDNA sequencing project".  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN (5)  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE-Bone marrow;  
 RA Zhao M., Song H., Li N., Peng Y., Han Z., Chen Z.,  
 RT "A novel gene expressed in human bone marrow".  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN (6)  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RA Howden P.,  
 RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Regulator of protein export for NES-containing proteins.  
 CC Also plays a role in mRNA nuclear export.  
 CC -1- SUBUNIT: Associates with NXF1, NXF2, NXF3 and NXF5.  
 CC -1- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and  
 CC the cytoplasm.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; A (shown here) and B: are  
 CC produced by alternative splicing.  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AJ277591; CAB96371.1;  
 DR EMBL: AJ278323; CAC01129.1;  
 DR EMBL: AF246127; AAF78034.1;  
 DR EMBL: AF201942; AAF86878.1;  
 DR EMBL: AK023289; BAB14511.1; ALT\_INT.  
 DR EMBL: AF212223; AAF87325.1; ALT\_INT.  
 DR EMBL: AL031387; CAB41301.1;  
 DR EMBL: AL031387; CAB41302.1;  
 DR MIM: 300320;  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2; 1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.

US-09-763-902B-10 (1-830) x NXL1\_DROME (1-133)

122 GATTTCAGACCTATGTCGATAGCGCTCAGACGCTGAGAGCATTTGTCATGCTAC 181

111 111 111111 111 111111 111



Db 4 AspleuLyAlaLyValGluSerCysAlaArgThrAlaSpHrPhrLeuTyr 23  
 QY 182 TACACACCATGATAGCGGCGCTTGTCTCCCGCTGATAGGACACGCCACC 241  
 Db 24 TyraLaseValaSpaHrArgGlnGlnIleGlyArgLeuTyrLeuSpaHrArg 43  
 QY 242 CTGGTCTGGAAGGCAATGCTTTGAGACAGAAATCTTGAGTAGTTTGAATG 301  
 Db 44 LeuSerTyrPaspngLyasnGlyAlaIleGlyArgGlnMetIleGluSerTyrPheGlnGlu 63  
 QY 302 TTGCCTTCAGGAGATTCACAGCGGTGAGTAGACGCGCTTCATGATGAGGAC 361  
 Db 64 LeuProSerSerAsnHisGlnLeuAsnThrLeuAspAlaGlnProIleValaspGlnAla 83  
 QY 362 ACACCAAGCAGACACCAAGCGTCTGTGTGATCATCTGTGATGAGTGTGAGGAGAC 421  
 Db 84 ValSerAsnGlnLeuAlaTyrLeuIleMetAlaSerGlySerValysPheAlaSpGln 103  
 QY 422 AAACACAGGAGCTTCAACCAAGATTCATCTGACCGCCAGCCGCTCACCAACACACA 481  
 Db 104 GlnLeuArgLyPheGlnGlnThrPheIleValThrAlaGlu-----AsnAsp 119  
 QY 482 GTGTGAGAGATCGAAGTACGTCGCTTCGCTCCAGGAC 520  
 Db 120 LysTrpLyValValSerAspCysTrpArgMetGlnGlu 132  
 RESULT 5  
 NNT1\_CAEEL STANDARD: PRT: 137 AA.  
 AC 09U757;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE NTF2-related export protein.  
 GN NNT1 OR Y719AM.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20036817; PubMed=10567585;  
 RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;  
 RT "Identification of an NTF2-related factor that binds Ran-GTP and  
 RT regulates nuclear protein export."  
 RL Mol. Cell. Biol. 19:8616-8624(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA Bradshaw-Cordum H., Scott K., Graves T.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Stimulator of protein export for NES-containing  
 CC proteins. Also plays a role in the nuclear export of. U1 snRNA,  
 CC tRNA, and mRNA (By similarity).  
 CC -1- SUBUNIT: Preferentially binds Ran-GTP (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 DR EMBL: AF156960; AAD54945.1;  
 DR EMBL: AC024852; AAK66028.1;  
 DR HSSP: P13662; 10UN.  
 DR WormPep: Y719AM.5; CE26780.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2.1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.

KW Transport; Protein transport; mRNA transport; Nuclear protein.  
 FT DOMAIN 19 135 NTF2.  
 SQ SEQUENCE 137 AA; 15479 MW; B99ABE3CA3479962 CRC64;  
 Alignment Scores:  
 Pred. No.: 1,39e-12 Length: 137  
 Score: 206.50 Matches: 40  
 Percent Similarity: 58.59% Conservative: 35  
 Best Local Similarity: 31.25% Identical: 48  
 Query Match: 13.87% Indels: 5  
 DB: 1 Gaps: 2  
 US-09-763-902b-10 (1-830) x NNT1\_CAEEL (1-137)  
 QY 140 GATCAGCGCTGCAGACCTCTGAGAGTTGTCAATGCTACTACACCATGATAG 199  
 Db 14 GlnGlnLeuGlyCysAsnGlnSerLysLysPheMetAspValTyrTrpAspAlaMetAspArg 33  
 QY 200 CGGCGCGCTTGTCTCCCGCTGATGATGAGGACAGCCACCTGCTGGAATGCAAT 259  
 Db 34 LysArgGlnLysIleGlyPheLeuTyrThrGlnValSerAsnAlaValTrpAsnGlyAsn 53  
 QY 260 GGTGTTTCAGACCAAGATCTTGAAGTATTTTGAATGTGCTTCACAGCAGTTC 319  
 Db 54 ProIleAsnGlyTyrAspSerIleCysGlnPheMetLyAlaLeuProSerThrGlnHis 73  
 QY 320 CAATCAGCGGTGATGACCTCCAGCCTGTCATGATGAAGCCACACCAAGCCAGACC--- 376  
 Db 74 AspIleGlnSerLeuAspAlaGlnArgLeuProGlnGlyValThrGlyAspMetSerGly 93  
 QY 377 ACGTCTTGTGTGTCATCTGTGATGAGTGTGAGGAGAACAAACAGCGGACTTC 436  
 Db 94 GlyMetLeuLeuAsnValAlaIleGlyAlaValThrValAspGlyAspSerLysArgAlaPhe 113  
 QY 437 AACCAAGACTTCATCTGACCGCCAGCGCTCACCCAGCAGACAGTGTGAAGATCGCA 496  
 Db 114 ThrGlnThrLeuLeuLeuGlyValGlu-----AspGlyLysTyrTrpValys 129  
 QY 497 AGTGAAGCTTCCGCTCCAGGAC 520  
 Db 130 SerAspArgPheArgTyrValasp 137  
 RESULT 6  
 NTF2\_HUMAN STANDARD: PRT: 127 AA.  
 ID NTF2\_HUMAN  
 AC P13662;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Nuclear transport factor 2 (NTF-2) (Placental protein 15) (PP15).  
 GN NTF2 OR PP15.  
 OS Homo sapiens (Human).  
 OS Mus musculus (Mouse).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606, 10090, 10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human.  
 RX MEDLINE=88247772; PubMed=3380696;  
 RA Grundmann U., Netlich C., Rein T., Lottspeich F., Kuemper H.A.;  
 RT "Isolation of cDNA coding for the placental protein 15 (PP15)."  
 RL Nucleic Acids Res. 16:4721-4721(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
 RC SPECIES-Human.  
 RX MEDLINE=95263674; PubMed=7744965;  
 RA Paschal B.M., Gerace L.;  
 RT "Identification of NTF2, a cytosolic factor for nuclear import that  
 RT interacts with nuclear pore complex protein p62."  
 RL J. Cell Biol. 129:925-937(1995).  
 RN [3]

```

DR EMBL: AK020512; BAB32122.1;
DR EMBL: U43939; AAA85905.1;
DR PIR: S00751; S00751.
DR PDB: 1O0N; 01-APR-97.
DR PDB: 1AR0; 24-DEC-97.
DR PDB: 1ASK; 28-JAN-98.
DR PDB: 1AZK; 29-APR-98.
DR MIM: 605813; -.
DR MGD: MGI:895065; Ntf2.
DR InterPro: IPR002075; NTF2.
DR Pfam: PF02136; NTF2.
DR PROSITE: PS50177; NTF2_DOMAIN; 1.
FW Transport: Protein transport; 3D-structure.
KT DOMAIN 10 121 NTF2.
SO SEQUENCE 127 AA; 14478 MW; 817752F20E262FD3 CRC64;

Alignment Scores:
Pred. NO.: 0.000234 Length: 127
Score: 124.00 Matches: 31
Percent Similarity: 43.59% Conservative: 20
Best Local Similarity: 26.50% Mismatches: 54
Query Match: 8.33% Indels: 12
DB: 1 Gaps: 3

US-09-763-902B-10 (1-830) x NTF2_HUMAN (1-127)

QY 167 TTTCGCAATGCTACTATACACCCACCTGATGAAGCGGGCGCTTGCTGTCGCCCGCTGTAC 226
Db ||||| ||||| ||||| |||||
QY 14 PheIleGlnHisTyrTyrGlnLeuPheAspAsnAspArgTglnLeuGlyAlaIleTyr 33
Db ||||| ||||| ||||| |||||
QY 227 ATGGGACACAGCCACCCCTGGTGTGAATGGCAATGCGTGTTCAGGACAGACATCCTTGAGT 286
Db ||||| ||||| ||||| |||||
QY 34 ILeAspAlaSerCysLeuThrTrpGlnGlyGlnGlnPheGlnGlnGlySalaIleVal 53
Db ||||| ||||| ||||| |||||
QY 287 GAGTGTGTGAAATGTGTCCTTCAGCGGATTCGCAATCGCGTGTA-----GACTGC 340
Db ||||| ||||| ||||| |||||
QY 54 GlyIysLeuSerSerLeuProheGlnGlyIleGlnHisSerLeuThrIleGlnAspHis 73
Db ||||| ||||| ||||| |||||
QY 341 CAGCGCTGTTCATGATGAAGCCACACCCAGCCAGACCGAGCCAGCGCTGTGTGTGATCTGTGA 400
Db ||||| ||||| ||||| |||||
QY 74 GlnProThrProAsp-----SerCysIleLeuSerMetValIleGly 87
Db ||||| ||||| ||||| |||||
QY 401 TCAGTGAAGTTGAGGGGCAACAACAACGCGGACTTAAACCAAGATCTCATCTGACCGCC 466
Db ||||| ||||| ||||| |||||
QY 88 GlnLeuIysAlaAspIleAspProIleMetGlyPheHisGlnMetPheLeu----- 105
Db ||||| ||||| ||||| |||||
QY 461 CAGCGCTCACCCGACACAGATGAGTGGAGATGGAGATGAGATGATCTCCGCC 511
Db ||||| ||||| ||||| |||||
Db 106 -----LysAsnIleAsnAspAlaTrpValCysThrAsnAspMetPheArg 120
Db ||||| ||||| ||||| |||||

RESULT 7
NTF2_XENLA STANDARD; PRT; 127 AA.
ID NTF2_XENLA 042424;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleot. transport factor 2 (Ntf2) (P10).
GN NTF2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBL_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RA "Pu R.T., Sheldis H.N., Dasso M.;
RT "Identification of a p10(Ntf2) homologue in Xenopus.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS.
CC INTERACTS WITH THE NUCLEOPORIN P62 AND WITH RAN. ACTS AT A
CC RELATIVELY LATE STAGE OF NUCLEAR PROTEIN IMPORT. SUBSEQUENT TO THE
CC INITIAL DOCKING OF NUCLEAR IMPORT LIGAND AT THE NUCLEAR ENVELOPE.

```

```
CC CC COULD BE PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).
CC CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sdb.ch/announce, or send an email to license@isb-sdb.ch).
CC CC -----
DR DR EMBL; AF023911; AAB81276.1; -
DR DR HSSP; P13662; IAKK.
DR DR InterPro; IPRO02075; NTF2.
DR DR Pfam; PF02136; NTF2_1.
DR DR PROSITE; PS50177; NTF2_DOMAIN; 1.
KW KW Transport; Protein transport.
FT FT DOMAIN 10 121 NTF2
SO SO SEQUENCE 127 AA; 14477 MW; 9711DD6828268259 CRC64;

Alignment Scores:
Pred. NO.: 0.000295 Length: 127
Score: 123.00 Matches: 31
Percent Similarity: 44.44% Conservative: 21
Best Local Similarity: 26.50% Mismatches: 53
Query Match: 8.26% Indels: 12
DB: Gaps: 3

US-09-763-902B-10 (1-830) x NTF2_XENLA (1-127)
QY 167 TTGTGCAATGCTACTACAGCACCAGTAGTAAGCGCGCGTGTCGCCGCTGTAC 226
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 14 PhellegingIntyrytGrInlthPhesaplaasparThngInleualValleyr 33
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 227 ATGGGCACAGCACCCCTGGTGTGAATGCGAATGCTGTTCAGGACAAAGAATCCTTGACT 286
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 34 ThraspalasSerCysLeuthrTyrrpGlnIylngInlyrhISglYalaAlallevela 53
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 287 GAGTTTTTTGAAATGTCCTTCAGCGAGTTCCAAATCAGCGGTGTA-----GACGTC 340
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 54 GlutylusLeusrleuleurProphengInlysIllegInhlSserIethrSergInasphIs 73
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 341 CAGCGCTGTTCATGATGAAGCCACACAGCAGACCGAGCGTCTGTTCATCTGTGTGA 400
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 74 GlnProthrProasp-----SerCysIlelleSerMetvalValgily 87
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 401 TCAATGAAAGTTTGAGGGGAGAACACACGCGGACTTCACCACTTATCTCTGACCGCC 460
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 88 GlnleulylsAlaaspaspAspproIleMetGlyPrhenISgInvalPhaleuleu----- 105
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 461 CAGCGCTTCACCGACACAACAGCATGTGGAGAATGCCAGATGAGTCTTCGCC 511
Db |||::||| ||||| ||| ||| ||| |||:::|||
QY 106 -----LysasnIllegInaspAlatrryValcystIrtnasnaspmetPhearg 120
Db |||::||| ||||| ||| ||| ||| |||:::|||

RESULT 8
NTF2_CANAL STANDARD: PRT: 124 AA.
AC AC Q9P926:
DT DT 16-OCT-2001 (Rel. 40, Created)
DT DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DE Nuclear transport factor 2 (Ntf-2).
GN GN NTF2.
OS OS Candida albicans (Yeast).
OC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX OX NCBI_Taxid=5476;
RN RN [1]
RP RP SEQUENCE FROM N.A. STRAIN-ATCC 26555;
RC RC
```

```

RA Bianchini-Roland S., Cordero-Otero R., Galliardin C., Herrero A.B.,
RA Dominguez A.,
RT "Sequence of Candida albicans and Yarrowia lipolytica homolog of the
RT Saccharomyces cerevisiae NTF2 gene."
RT Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE
CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (By SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF145758; AAF66701.1;
DR HSSP; P13662; 10UN.
DR InterPro; IPR002075; NTF2.
DR Pfam; PF02136; NTF2; 1.
DR PROSITE; PS50177; NTF2_DOMAIN; 1.
KW Transport; Protein transport.
FT DOMAIN
FT SEQUENCE 124 AA; 14193 MW; B799D1ABA828A049 CRC64;
SQ
Alignment Scores:
Pred. No.: 0 000825 Length: 124
Score: 118.50 Matches: 40
Percent Similarity: 42.65% Conservative: 18
Best Local Similarity: 29.41% Mismatches: 57
Query Match: 7.96% Indels: 21
Gaps: 6
US-09-763-902B-10 (1-830) x NTF2_CANAL (1-124)
QY 116 TCTGTGGATTTCAAGACCTATGTGATCGACGCGCTCGAGAGCTGTGAGAGTGTGCAAT 175
Db 2 SerValAspPheHisnla-----ValAlaThrGlnPheCysAsn 14
QY 176 GTCTACTACACCAACCATGATTAAGCGCGCGCTTGTCTGTCCGCGCTGTACATGGGCACA 235
Db 15 PheTYrTYfAsnGlnPheAspSerAsparGserGlnLeuGlyAsnLeuTYrArgAsnGlu 34
QY 236 GCACACCGCTGTGGAAATGGCAATGCTGTTCAGGACCAAGAACTCCTGACGTATTTT 295
Db 35 SerMetLeuThrPheGlnIuThrSerGlnLeuGlnGlyAlaIarGAspIleValGlnLYsLeu 54
QY 296 GAAATGTTCGCTTCACGACGAG-----TTCCAAATCAGCGTGTGAGACTGCCACCGCTGTT 349
Db 55 AlaSerLeuProPheGlnLYsValAlaHisArgIleSerThrIleAspAlaGlnPro--- 73
QY 350 CATGATGAAGCCACACCAAGCCACACGACGCGCTGTTCGATCTGTGATCA----- 403
Db 74 -----AlaSerAlaAsnGlyAspIleLeuValMetValThrGlyGluLeu 89
QY 404 GATGAGTTTGAGGGGGAACAACAAGGAGGAGCTTCACACCAAGAACTTCATCCTGACCGCCAG 465
Db 90 IleAspGlnGlnGlnAsnAlaGlnIarG--TYrSerGlnValPheHisLeuIle----- 106
QY 464 GCCTCACCGACGACACACAGTGTGGAAGATCGCAAGTGAAGTACTGCTTCGCC 511
Db 107 -----ProAspAsnGlySerTYrTYrValPheAsnAspIlePheArg 120
RESULT 9
NTF2_ARATH
ID NTF2_ARATH STANDARD: PRT: 126 AA.
AC Q9C7F5;
FT 16-OCT-2001 (Rel. 40, Created)
FT 16-OCT-2001 (Rel. 40, Last sequence update)
FT 16-OCT-2001 (Rel. 40, Last annotation update)

```



DR HSPB, P13662; 100N.  
 DR SCD, S0000811; NTF2.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF01316; NTF2; 1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
 DR Transports: Protein transport.  
 FT DOMAIN 8 121 NTF2.  
 SO SEQUENCE 125 AA; 14453 MW; 56608C00E634714 CRC64;

## Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	0.00131	125	42	19	54	21
Percent Similarity:	116.50					
Best Local Similarity:	44.85%					
Query Match:	7.82%					

US-09-763-902b-10 (1-830) x NTF2\_YEAST (1-125)

OY 116 TCTGTGATTCAGACCTATGTGATCAGCGCTTCAGAGCTTGTCAAT 175  
 DB 2 SerLeuaspPheasnThr-----LeuAlaGlnasnPheThrgln 14  
 OY 176 GTCTACTACACCATGATGATAGCGCGCTTGTCTGCCGCTGTACATGGCACA 235  
 DB 15 PheTyrTyrAsnGlnPheAspPheThrsparserGlnLeuGlyAsnLeuTyrArgAsnGln 34  
 OY 236 GCCACCTCGTCTGGAATGCAATGCTCTTCAGAGCAAGAACTCTTGAGTATTTT 295  
 DB 35 SerMetLeuThrPheGlnThrSerGlnLeuGlnGlyAlaTyrAspPheLeuGlyAsnLeu 54  
 OY 296 GAAATGTGCTCTCCAGAGTTCACAA-----ATCAGCGTGTGACTGACCGAGCTGT 349  
 DB 55 ValSerLeuPhePheGlnGlnValGlnHisArgIleThrLeuAspAlaGlnPro--- 73  
 OY 350 CATGATGAGACCAACCAACGACGACGCGCTTGTCTGTGCTGATGATGATCA----- 403  
 DB 74 -----AlaSerProasn---GlyAspValLeuValMetIleThrGlyAspLeuLeu 89  
 OY 404 GTGAAGTTTGAAGGGAGCAACACGAGCTTCACCAAGCTTATCTGACCGCCAG 463  
 DB 90 IleaspGlnGlnGlnAsnProGlnArg---PheSerGlnValPheHisLeuIle----- 106  
 OY 464 GCCTCACCCAGCAACACAGTGTGGAAGATCGCAATGACTGCTTCCGC 511  
 DB 107 -----ProaspGlyAsnSerTyrTyrValPheAsnAspIlePheArg 120

RESULT 11  
 G3B2\_HUMAN STANDARD: PRT: 482 AA.  
 AC Q9UN86; O75149; O60606; Q90PA1;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 15-JUN-2002 (Rel. 41; Last annotation update)  
 DE Ras GTPase-activating protein binding protein 2 (GAP SH3-domain binding protein 2) (G3BP-2).  
 GN G3BP2 OR KIA0660.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid:9606;  
 RN [1]  
 RN SEQUENCE FROM N.A. (ISOFORMS A AND B).  
 RC TISSUE-Brain;  
 RA Kennedy D., Mattick J.S.;  
 RT "Characterisation and chromosomal location of G3BP-1 and G3BP-2a/b, members of a novel SH3 domain-binding and RNA-binding protein family implicated in signal transduction";  
 RT Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A. (ISOFORM A).  
 RC TISSUE-Brain;  
 RX MEDLINE:98403880; PubMed:9734811;

RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";  
 RT DNA Res. 5:169-176(1998).  
 RN [3]  
 RN SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE-Brain;  
 RA Gultard E.;  
 RT Submitted (Feb-1998) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RN SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE-Brain;  
 RA Kennedy D., Ru K., Mattick J.S.;  
 RT Submitted (Mar-1998) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RN SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE-B-cell;  
 RA Strausberg R.;  
 RT Submitted (Jul-2001) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RN INTERACTION WITH IKAPPABALPHA.  
 RX MEDLINE:20549669; PubMed:10969074;  
 RA Prigent M., Barlat I., Langen H., Dargemont C.;  
 RT "Ikappabalpha and Ikappabalpha/NF-kappa B complexes are retained in the cytoplasm through interaction with a novel partner, RasGAP SH3-binding protein 2.";  
 RT J. Biol. Chem. 275:36441-36449(2000).  
 CC -1- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA TRANSPORT (POTENTIAL).  
 CC -1- SUBUNIT: BINDS TO THE N-TERMINAL DOMAIN OF IKAPPABALPHA.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF145284; AAD51932.1; -;  
 CC EMBL: AB014560; BAA31635.1; -;  
 CC EMBL: AF051311; AAC15705.1; -;  
 CC EMBL: AF053535; AAC95292.1; -;  
 CC EMBL: BC011731; AAH11731.1; -;  
 CC HSPB, P09651; IHA1.  
 DR InterPro: IPR002075; NTF2.  
 DR InterPro: IPR000504; RNA\_rec\_mot.  
 DR Pfam: PF00076; rtm; 1.  
 DR Pfam: PF02136; NTF2; 1.  
 DR SMART: SM00360; RRM; 1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
 DR PROSITE: PS50102; RRM; 1.  
 DR PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 RT Transport; RNA-binding; Alternative splicing.  
 FT DOMAIN 11 133 NTF2.  
 FT DOMAIN 134 409 RNA-BINDING (RRM).  
 FT DOMAIN 134 223 GLY-RICH.  
 FT VARSPIC 243 479 MISSING (IN ISOFORM B).  
 FT CONFLICT 267 267 S -> P (IN REF. 2).  
 FT CONFLICT 359 359 E -> V (IN REF. 3).  
 FT CONFLICT 460 460 M -> I (IN REF. 3).  
 SO SEQUENCE 482 AA; 54111 MW; 254506A3F1AAE218 CRC64;

## Alignment Scores:

Pred. No.:	Length:	Matches:
0.00226	482	

Score: 115.50 Matches: 38  
 Percent Similarity: 44.09% Conservative: 18  
 Best Local Similarity: 29.92% Mismatches: 58  
 Query Match: 7.76% Indels: 13  
 DB: 1 Gaps: 4

US-09-763-902b-10 (1-830) x G3B2\_HUMAN (1-482)

OY 164 GAGTTTGCAATGCTACTACACCAACCATGATAGCGGCGGCGTTCGTCGCCGCTG 223  
 DB 14 GIupheValaIrgLInTyTThLeuLeuAnLysAlaProGIuTyLeuHisArgPhe 33  
 OY 224 TACATGGGACAGCCACCGCTGTCGTAATGCG-----AAT 259  
 DB 34 TyrgIyArgAnSerSerTyValHisGIyGIyValaIAsPalaSerGIyLysProGIInGIu 53  
 OY 260 GCTGTTTACAGACAAAGATCCTTGAGTGAGTTTGAATGCTTCCTCCAGCAGATTC 319  
 DB 54 AlavaIlyrGIyGIInAsnAspIleHisIlySValLeuSerLeuAnPheSerGIuLys 73  
 OY 320 CAATCAGCGTGTGAGTGCACCGCTTCATGATGAGCCACCAAGCCAGACGACG 379  
 DB 74 HIsThrLysIle-----ArgHisValaIAsPalaHisAlaThLeuSerAsp---GIy 89  
 OY 380 GTCTGTGTGTCATCTGTGATCAGTGAAGTTTGAGGGGAACAACAACGCGACTTCAC 439  
 DB 90 ValValaIValGIInValMetGIyLeuLeuSerAnSerGIyGIInProGIuArgLysPheMet 109  
 OY 440 CAGAACTTCATCCTACCGCCGCGCTCACCCACCAACAGTCGTGAAGATCCCAAGT 499  
 DB 110 GIInThrPheValLeuAlaProGIuLysSerValProAnLysPheTy---ValHisAsn 128  
 OY 500 GACTGCTTCGCGCTCCAGGAC 520  
 DB 129 AspMetPheArgTyGIuLasp 135

## RESULT 12

G3B2\_MOUSE STANDARD: PRT: 482 AA.

ID G3B2\_MOUSE STANDARD: PRT: 482 AA.

AC 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ras-GTPase-activating protein binding protein 2 (GAP SH3-domain

binding protein 2) (G3BP-2).

GN G3BP2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM B).

RX MEDLINE=9826259; PubMed=9575347;

RA Kennedy D., Wood S.A., Ramsdale T., Tam P.P., Steiner K.A.,

RA Mattick J.S.;

RT Identification of a mouse orthologue of the human ras-GAP-SH3-domain

RT binding protein and structural confirmation that these proteins

RT contain an RNA recognition motif."

RL Biomed. Pept. Proteins Nucleic Acids 2:93-99(1997).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS A AND B).

RA Kennedy D., Mattick J.S.;

RT Characterisation and chromosomal location of G3BP-1 and G3BP-2a/b,

RT members of a novel SH3 domain-binding and RNA-binding protein family

RT implicated in signal transduction."

RL Submitted (APR-1999) to the EMBL/GenBank/DBI databases.

CC -1- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN RNA

CC TRANSPORT (POTENTIAL).

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A (SHOWN HERE) AND B; ARE

CC PRODUCED BY ALTERNATIVE SPLICING.

CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isp-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).

CC EMBL: U65313; AAC53553.1;  
 DR EMBL: AF145285; AAD51933.1;  
 DR HSSP: P09651; 1HA1; G3BP2.  
 DR MGD: MGI:1346339; G3BP2.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF00076; rrm; 1.  
 DR Pfam: PF02136; NTF2; 1.  
 DR SMART: SM00360; RRM; 1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
 DR PROSITE: PS50102; RRM; 1.  
 DR PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 DR Transports: RNA-binding; Alternative splicing.  
 KW DOMAIN 11 133  
 FT DOMAIN 331 409  
 FT DOMAIN 134 223  
 FT DOMAIN 419 479  
 FT VARSPLIC 243 275  
 FT CONFLICT 476 482  
 FT MISSING (IN ISOFORM B).  
 FT REFGOR -> TLHRTASLKSQWQFMOWYTLHRVCILWKF  
 FT FLVMTQTF (IN REF. 1).  
 SO SEQUENCE 482 AA; 54088 MW; F8BBBC2086123A5CE CRC64;

Alignment Scores:  
 Pred. No.: 0.00226 Length: 482  
 Score: 115.50 Matches: 38  
 Percent Similarity: 44.09% Conservative: 18  
 Best Local Similarity: 29.92% Mismatches: 58  
 Query Match: 7.76% Indels: 13  
 DB: 1 Gaps: 4

US-09-763-902b-10 (1-830) x G3B2\_MOUSE (1-482)

OY 164 GAGTTTGCAATGCTACTACACCAACCATGATAGCGGCGGCGTTCGTCGCCGCTG 223

DB 14 GIupheValaIrgLInTyTThLeuLeuAnLysAlaProGIuTyLeuHisArgPhe 33

OY 224 TACATGGGACAGCCACCGCTGTCGTAATGCG-----AAT 259

DB 34 TyrgIyArgAnSerSerTyValHisGIyGIyValaIAsPalaSerGIyLysProGIInGIu 53

OY 260 GCTGTTTACAGACAAAGATCCTTGAGTGAGTTTGAATGCTTCCTCCAGCAGATTC 319

DB 54 AlavaIlyrGIyGIInAsnAspIleHisIlySValLeuSerLeuAnPheSerGIuLys 73

OY 320 CAATCAGCGTGTGAGTGCACCGCTTCATGATGAGCCACCAAGCCAGACGACG 379

DB 74 HIsThrLysIle-----ArgHisValaIAsPalaHisAlaThLeuSerAsp---GIy 89

OY 380 GTCTGTGTGTCATCTGTGATCAGTGAAGTTTGAGGGGAACAACAACGCGACTTCAC 439

DB 90 ValValaIValGIInValMetGIyLeuLeuSerAnSerGIyGIInProGIuArgLysPheMet 109

OY 440 CAGAACTTCATCCTACCGCCGCGCTCACCCACCAACAGTCGTGAAGATCCCAAGT 499

DB 110 GIInThrPheValLeuAlaProGIuLysSerValProAnLysPheTy---ValHisAsn 128

OY 500 GACTGCTTCGCGCTCCAGGAC 520

DB 129 AspMetPheArgTyGIuLasp 135

RESULT 13

NTF2\_ORYSA STANDARD: PRT: 122 AA.

ID NTF2\_ORYSA STANDARD: PRT: 122 AA.

AC 09XJ54; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Nuclear transport factor 2 (Ntf2-2).  
GN NTF2.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Erihartoideae; Oryzaceae; Oryza.  
ON NCBI\_Taxid=4530;  
RN [1]  
RP SEQUENCE FROM N. A.  
RC STRAIN=CV. Nipponbare;  
RA Matsui R., Iwasaki T., Jiang C., Yamamoto N.;  
RT "Molecular cloning of a cDNA encoding nuclear transport factor 2  
(Ntf2) from rice."  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE  
CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE  
CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AB01162; BAA81910.1;  
DR HSSP: P13662; 1A2K.  
DR InterPro: IPR002075; NTF2.  
DR Pfam: PF02136; NTF2.  
DR PROSITE: PS50177; NTF2\_DOMAIN. 1.  
DR KEGG: K01199; NTF2.  
FT DOMAIN 6 119 NTF2.  
FT SEQUENCE 122 AA; 13354 MW; 84100EBB62AB1BD1 CRC64;  
SQ  
Alignment Scores:  
Pred. No.: 0.016 Length: 122  
Score: 107.00 Matches: 35  
Percent Similarity: 46.28% Conservative: 21  
Best Local Similarity: 28.93% Mismatches: 53  
Query Match: 7.19% Indels: 12  
Gaps: 4  
US-09-763-902b-10 (1-830) x NTF2\_ORYSA (1-122)  
QY 158 GCGAGGAGTTGTCATGCTCTACTACACCAATGATAGCGCGCTTCTGCTCC 217  
DB 7 AlAlvysAlaPhelAlGluHsYrYrYrPhesPhrAsnaAProAlaLeuVal 26  
QY 218 CGCCTGATCATGGGACACCGCTGCTGATGCAATGCTGTTTTCAGGACAGAA 277  
DB 27 SerLeuTyrlnspsglYserMetLeuThrPhelInclnglnPheLeuGlyAlaAla 46  
QY 278 TCCTTAGAGTAGTTTGAATGCTGCTTCAGCAG-----TTCCAATCAGCGTG 331  
DB 47 AlAlleAlaGlySerLeuGlySerLeuProPhelInclnglnShsHsAlleAsrThr 66  
QY 332 GTGAGCTGCACAGCTGTTCATGATGAAGACACCAACGACAGCGCTTGTGTC 391  
DB 67 ValAspCysGlnPro-----SerGlyProGlnGlyMetLeuValPhe 81  
QY 392 ATCTGGAGTCACTGAAGTTTGAAGGAGAACAAACA-----CGGAGCTCAACCGAATGTC 448  
DB 82 ValSerGlySerLeuTyrlnglyProAspGlnHisProLeuTyrlsPheSerGlnMetPhe 101  
QY 449 ATCTCAACGCGCCAGCGCTTCACCGACACACAGTGTGAGAGTGCAGAGTGCCTTC 508  
DB 102 -----GlnLeuLeuProAlaGlyGlyAsnPheryValGlnAsnAspMetPhe 117

QY 509 CGC 511.  
DB 118 Arg 118  
RESULT 14  
NTF2\_SCHPO  
ID NTF2\_SCHPO STANDARD; PRT; 123 AA.  
AC 010100; OSUTF4;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable nuclear transport factor 2 (Ntf2-2).  
GN NTF2 OR SPAC1B9.01C OR SPAC15P9.03C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomyces.  
ON NCBI\_Taxid=4896;  
RN [1]  
RP SEQUENCE FROM N. A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
RA Jones K., Jones M., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moulé S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Voiclaert G., Aert R., Robben J., Gymnopoulos B.,  
RA Weltjens I., Vanstreels E., Rieger M., Scheeler M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzjak K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Motier S.,  
RA Gallibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito G.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe."  
RL Nature 415:871-880(2002).  
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE  
CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE  
CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AL109931; CAB53052.1;  
DR HSSP: 268197; CAA92380.2;  
DR InterPro: IPR002075; NTF2.  
DR Pfam: PF02136; NTF2.  
DR PROSITE: PS50177; NTF2\_DOMAIN. 1.  
DR KEGG: K01199; NTF2.  
FT DOMAIN 7 120 NTF2.  
FT SEQUENCE 123 AA; 14078 MW; F587303CA9270E6 CRC64;  
SQ

Alignment Scores: 0.0116 Length: 123  
 Pred. No.: 107.00 Matches: 35  
 Score: 46.288 Conservative: 21  
 Percent Similarity: 28.93% Mismatches: 53  
 Best Local Similarity: 28.93% Indels: 12  
 Query Match: 7.19% Gaps: 4

US-09-763-902b-10 (1-830) x NTF2\_SCHPO (1-123)

OY 158 GCTGAGAGTTTGTCAATGCTCTACTACACACCATGATAAGCGCGCTTCTCTCC 217  
 Db 8 AAlathrGlnphetrGlnphetrGlnphetrGlnphetrGlnphetrGlnphetr 27  
 OY 218 CGCTGTACATGGCAGACACCGCTGCTGGAATGCTGTTTTCAGGACAGAA 277  
 Db 28 SerLeuYrArGlnGlnSerMetLeuSerMetLeuGlnGlnGlnGlnGlnGln 47  
 OY 278 TCCTTGAGTGAAGTTTGAATGCTCTCCAGCGAGTTCCAA-----ATCAGCGTG 331  
 Db 48 AAlaIleValGlnValSerLeuProheGlnArGlnHisArgIleSerThr 67  
 OY 332 GTACACTGCCAGCGCTGTATGATGATGACACACACCGACCGCTCTGTTGTC 391  
 Db 68 LeuAspAlaGlnPro-----ThrglyThrThrGlySerValIleValMet 82  
 OY 392 ATCTGTGATCATGAGTATT---GAGGGGAAACAACAGCGGACTTCACAGACTTC 448  
 Db 83 ValThrGlyGlnLeuLeuAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 102  
 OY 449 ATCTGACCGCGCCAGCGCTCACCACACAGAGTGAAGTGAAGTGAAGTGAAGT 508  
 Db 103 HisLeuVal-----AsnAsnAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 118

OY 509 CGC 511  
 Db 119 Arg 119

RESULT 15  
 G3BP\_HUMAN STANDARD: PRT: 466 AA.  
 ID 013283;  
 AC 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ras-GTPase-activating protein binding protein 1 (GAP SH3-domain binding protein 1) (G3BP-1).  
 GN G3BP OR G3BP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RA Parker F., Maurier F., Delumeau I., Duchesne M., Faucher D., Debussche L., Dugue A., Schweighoffer F., Tocque B.,  
 RT "A Ras-GTPase-activating protein SH3-domain-binding protein."  
 RL Mol. Cell. Biol. 16:2561-2569(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Prostate;  
 RA Strausberg R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA TRANSPORT (POTENTIAL).  
 CC -1- SUBUNIT: BINDS TO THE SH3 DOMAIN OF RAS-GTPASE-ACTIVATING PROTEIN (GAP120).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS.  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/ebis/ebis-sib.ch/announce/>  
 CC or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC -----  
 DR EMBL: U32519; AAB07787.1;  
 DR EMBL: BC006997; AA06997.1;  
 DR HSPB; P09651; IHA1;  
 DR InterPro: IPR002075; NTF2.  
 DR InterPro: IPR000504; RNA\_rec\_mot.  
 DR Pfam: PF00076; rrm; 1;  
 DR Pfam: PF02136; NTF2; 1;  
 DR SMART: SM00360; RRM; 1;  
 DR PROSITE: PS01177; NTF2\_DOMAIN; 1.  
 DR PROSITE: PS01012; RRM; 1; FALSE\_NEG.  
 DR PROSITE: PS00030; RRM\_RNP; 1;  
 KW Transport; RNA-binding.  
 FT DOMAIN 11 133 NTF2.  
 FT DOMAIN 340 415 RNA-BINDING (RRM).  
 FT DOMAIN 144 225 GLU-RICH.  
 FT DOMAIN 430 461 GLY-RICH.  
 SQ SEQUENCE 466 AA; 52164 MW; 0F9429D78E0C7F59 CRC64;

Alignment Scores: 0.0223 Length: 466  
 Pred. No.: 105.50 Matches: 36  
 Score: 44.868 Conservative: 21  
 Percent Similarity: 28.35% Mismatches: 57  
 Best Local Similarity: 28.35% Indels: 13  
 Query Match: 7.09% Gaps: 4

US-09-763-902b-10 (1-830) x G3BP\_HUMAN (1-466)

OY 164 GAGTTTGTCAATGCTCTACTACACACCATGATAAGCGCGCTTCTCTCCGCTG 223  
 Db 14 GluPheValArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 33  
 OY 224 TACATGGGACACGCGCTGCTGGAATGCG-----AAT 259  
 Db 34 TyrGlyLysAsnSerSerTyrValHisGlyLeuAspSerAsnGlyLysProAlaAsp 53  
 OY 260 GCTGTTTCAGACAGAAATCCTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 319  
 Db 54 AlaValTyrGlyGlnLys-----GluIleHisArgLysValMetSerGlnAsnPro 70  
 OY 320 CAATCAGCGGTGTAGACTGCCAGCTTTCATGATGATGATGATGATGATGATGAT 379  
 Db 71 ThrAsnGlnHisThrLysIleArgHisValAspAlaHisAlaThrLeuAsnAsp---Gly 89  
 OY 380 GTCCTGTGTGATCATGTGTGATCATGTGATCATGTGATCATGTGATCATGTGAT 439  
 Db 90 ValValValGlnValMetGlyLeuLeuSerAsnAsnGlnAlaLeuValArgPheMet 109  
 OY 440 CAGAACTCATCTGACCGCCGACGCGCTCACCACGACGACGACGACGACGACGAC 499  
 Db 110 GlnThrPheValLeuValProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 128  
 OY 500 GACTGCTTCCGCTCCAGCAGC 520  
 Db 129 AspIlePheArgTyrGlnAsp 135

Search completed: June 24, 2003, 19:32:09  
 Job time : 16.6012 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 19:25:38 ; Search time 57.1255 Seconds

(without alignments)  
5987:487 Million cell updates/sec

Title: US-09-763-902b-10

Perfect score: 1489

Sequence: 1 tgaacttgatcactgctg.....ataattataatacactg 830

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+12p\_model -DEV=xip  
-O=/cgn2.1/USPO/USO9763902/unat\_24062003\_130341\_12453/APP-query.fasta.1.2254  
-DB=SPTRMBL\_21 -OPMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -STARR=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pro -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=USO9763902 -ECGN\_1.1.232 -efunat\_24062003\_130341\_12453 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEOQUERY -NEG\_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

SPTRMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriopl:\*  
17: sp\_archaeopl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155.5	10.4	167	5	Q8T2G7 dictyostell

2	142	9.5	200	3	Q9HDY6	Q9hdy6 schizosach
3	135.5	9.1	146	10	Q8RUJ2	Q8ruj2 oryza sativ
4	129	8.7	127	5	Q8R2D5	Q8r2d5 dictyostell
5	125.5	8.4	123	3	Q9P8H0	Q9p8h0 yarrowia ll
6	119	8.0	688	5	Q9I7J5	Q9i7j5 drosophila
7	119	8.0	690	5	Q9NH72	Q9nh72 drosophila
8	119	8.0	690	5	Q9VET4	Q9vet4 drosophila
9	118	7.9	522	10	Q9LPY0	Q9lpy0 arabidopsis
10	112.5	7.6	125	3	Q9AVN3	Q9avn3 emericella
11	106	7.1	537	5	Q2I351	Q2i351 caenorhabd
12	105.5	7.1	130	5	Q9VJ85	Q9vj85 drosophila
13	105	7.1	122	10	Q9FZK4	Q9fzk4 arabidopsis
14	103.5	7.0	796	3	Q9HE57	Q9he57 kluyveromy
15	102	6.9	218	4	Q6O599	Q6o599 homo sapien
16	99	6.6	966	10	Q8S0Z3	Q8s0z3 oryza sativ
17	96	6.4	903	4	Q9UPX1	Q9upx1 homo sapien
18	93	6.2	450	10	Q9FNDO	Q9fndo arabidopsis
19	92.5	6.2	590	10	Q9LD99	Q9ld99 oryza sativ
20	91	6.1	285	12	Q9VYP6	Q9vyp6 avian adeno
21	90.5	6.1	666	3	Q8TFR8	Q8tfr8 saccharomyc
22	90.5	6.1	666	3	Q8TFR7	Q8tfr7 saccharomyc
23	90.5	6.1	685	3	Q9HGV1	Q9hgv1 saccharomyc
24	90.5	6.1	685	3	Q8TFB8	Q8tfb8 saccharomyc
25	90	6.0	472	11	Q9OX13	Q9ox13 mus musculu
26	88.5	5.9	129	4	Q9HAF6	Q9haf6 homo sapien
27	88.5	5.9	130	5	Q9VRD6	Q9vrd6 drosophila
28	88.5	5.9	353	17	Q9Y998	Q9y998 aeropyrum p
29	88.5	5.9	507	4	Q9Y998	Q9y998 homo sapien
30	88.5	5.9	4823	13	Q93321	Q93321 fuigu rubrip
31	87.5	5.9	460	10	Q9FME2	Q9fme2 arabidopsis
32	87.5	5.9	629	12	Q9DWH3	Q9dwh3 rat cytomeg
33	87	5.8	485	10	Q9LXL1	Q9lxl1 arabidopsis
34	86.5	5.8	563	13	Q9DDJ4	Q9ddj4 halichoeres
35	86.5	5.8	719	10	Q8S7D0	Q8s7d0 oryza sativ
36	86.5	5.8	109	10	Q64428	Q64428 chorella v
37	86	5.8	7662	4	Q10465	Q10465 homo sapien
38	86	5.8	34350	4	Q8W242	Q8w242 homo sapien
39	85.5	5.7	147	4	Q9P058	Q9p058 homo sapien
40	85.5	5.7	753	11	Q9JHW8	Q9jhw8 ratian novr
41	84.5	5.7	872	15	Q9O0Y5	Q9o0y5 human immun
42	84.5	5.7	926	16	Q92L26	Q92l26 rhizobium m
43	84	5.6	357	6	Q97866	Q97866 sus scrofa
44	84	5.6	362	4	Q9BRD4	Q9brd4 homo sapien
45	84	5.6	499	4	Q8TCP7	Q8tcp7 homo sapien

## ALIGNMENTS

### RESULT 1

Q8T2G7 PRELIMINARY; PRT; 167 AA.

AC Q8T2G7; 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical 19.0 kDa protein.  
OS Dictyostellium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium;  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RX4;  
RA Gloeckner G., Eichinger L., Szafianski K., Pachbat J., Dear P.,  
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noejei A.A.,  
RT "Sequence and Analysis of Chromosome 2 of Dictyostellium".  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBD databases.  
DR EMBL; AC115607; AAL92371.1;  
KW Hypothetical protein.  
SQ SEQUENCE 167 AA; 19039 MW; E16E4A5F4DA0DBAA CRC64;

Alignment Scores: 6.47e-08 Length: 167  
Pred. No.: 1

Score:	155.50	Matches:	21
Percent Similarity:	45.39%	Conservative:	43
Best Local Similarity:	29.08%	Mismatches:	68
Query Match:	10.44%	Indels:	9
DB:	5	Gaps:	5

US-09-763-902B-10 (1-830) x Q9HDY6 (1-167)

QY	104	CCAGAGATGGCATCTGTGGAT-----TTCAAGACTATGTGGATTCAGGCGCTGCAGAGCT	157
	111	:::	111
Db	28	ProthrasnglyasnyvalasplalalalsleuylsilevalglutnrllethrProarg	47
QY	158	GCAGAGAGATTGTGCT---AATGCTACATACACACACATGATTAAGCGCGCGCTTGCTG	214
Db	48	AleghlvalphellelysgludherlytrProlystyAspserSerthgalaAspleu	67
QY	215	TCCGCGCTGTACATGGGACAGCCAGCCCTGGTGTGGAATGGCAATGCTGTTTCAGGCA	274
		:::	
Db	68	IlleghlyleuyltylasyasphlsserValserlletfrpangltylthglucylsglyPro	87
QY	275	GAATCTGTAGAGATGTTTTTAAATGTTGCTTCACAGCGAGTTCCAAATCAGCGTGTGA	334
		:::	:::
Db	88	GLnHlsIleolylysleuLeuualagluileProAsnSerValnHsvalvalglutnhrPhe	107
QY	335	GACTGGCAGCGCTGT-----CATGTGTAAGCCACACACAGCAGCAGCAGCGCTGTGT	388
Db	108	AsplalglInProvalProSerAspshlysglAsnProasn-----lleuile	124
QY	389	GTCATCTGTGTGATCACTGAATGTTTGAGGGGAACAACAAGCGGACTTCACACAGACTTC	448
Db	125	ThralatTheglyylstvalthryltylthtrhserglInHlsglInPhnHlsglInthhrPhe	144
QY	449	ATCCTGACGCGCCAGCGCTCACACCAGCAACAGTGTGGAAAGATGGCAAGTACTGCTTC	508
	:::	:::	:::
Db	145	LeuLeuVallylAspProthrasnSerAsn--LeuPhetylLeuSerTyAspCysIle	163
QY	509	CGC 511	
Db	164	Arg 164	

RESULT 2

Q9HDY6 PRELIMINARY; PRT: 200 AA.

AC Q9HDY6: 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Putative nuclear export factor.

GN SPAPB1A10.03.

OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

CC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID:4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN:972H-;

RA O'Neil S., Harris D., Wood V., Rajandream M.A., Barrrell B.G.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL512562; CAC1476.1;

DR InterPro: IPR002075; NTF2.

DR Pfam: PF02136; NTF2.1.

SQ SEQUENCE 200 AA: 23029 MW: B05E47D62B2B7BF CRC64;

Alignment Scores:

Pred. No.:	1,96e-06	Length:	200
Score:	142.00	Matches:	31
Percent Similarity:	52.89%	Conservative:	33
Best Local Similarity:	25.62%	Mismatches:	43
Query Match:	9.54%	Indels:	14
DB:	3	Gaps:	4

US-09-763-902B-10 (1-830) x Q9HDY6 (1-200)

OY	137	GTGATACAGCGCTCAGACGTCGTGAGCAAGTTTGCATGCTACTACACCACATGGAT	196
Db	1	MelJuserSerValLysTrpAlaGlnIleupheValGlnAlygTyTrpSerSerLeuasp	20
OY	197	AAGCGGCGGCTTGCTGCCCGCTTACATGAGGGCACAGCCACCCTGGTGTGAATGGC	256
Db	21	ThrsnArghangLyIlealagIunpheyrArngIuanserLeuIIeleurpAsngly	40
OY	257	AATGCTGTTTCAGACAAGAATCCTTGAGTGAGTTTTTGAATG-----TTGCTT	307
Db	41	LysPromet-----GlnValThrGluPheThrSerMetIleValasnLeupro	56
OY	308	TCCAGCGAGTTCCAATACAGCGGTGTAGACCTCCAGCGCTTCATCATGATGAACCCACCA	367
Db	57	TySerLysTrpHrLysValGluAspPheaspsrgInglinal-----MetGly	72
OY	368	AGCCAGACACCGGCTCCTGTTGTATCATCTGTGATCAGTGAAGATTGTAGGGGAACAACA	427
Db	73	AsnAspMetalaInlleilelvalValserGlyThrIleArgPheaspolyLysTrpPro	92
OY	428	CGGAGCTTCACACCACTTCATCTGTACCGCGCCAGCGCTCACCCAGCAACAACAGGTGG	487
Db	93	HlsValPheSerTyArg-----SerAlaThrTrpGlnProkspAspThrIlepe	109
OY	488	AAG	490
Db	110	Lys	110

RESULT 3  
ID OBRU32 PRELIMINARY; PRT; 146 AA.  
AC OBRU32:  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DE Putative nuclear transport factor 2.  
GN P0415A04.24.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae;  
OC Ehrhartoideae: Oryzeae; Oryza.  
OX NCBI\_TaxID=39947;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC  
clone:P0415A04."  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBSJ databases.  
DR EMBL; AP003345; BAB90110.1; -  
SQ SEQUENCE 146 AA; 15796 MW; 496BE9A02EIE9670 CRC64;

Alignment Scores:

Pred. No.:	9,25e+06	Length:	146
Score:	135.50	Matches:	45
Percent Similarity:	48.15%	Conservative:	20
Best Local Similarity:	33.33%	Mismatches:	54
Query Match:	9,10%	Indels:	16
	10	Gaps:	6

US-09-763-902B-10 (1-830) x OBRU32 (1-146)

OY	121	GGATTTCAGACCTAATGTGATCAGGCTCCTCAGACGCTGCTGAGAGTTTGTCAATGCTTA	180
Db	20	GlyTrpSerAspGlnCys-AspValValAlaArgAla-----PheValGluTrpTy	36
OY	181	CTACACCAACCATGATAAAGCGGCGGCTTGCTGTCTCCCGCTTACATGGGCAACGCAC	240
Db	36	TyrcGlnThrPheasPrpHrasnArghAlaIaleuAlaIaleuTyrcIyGlnThrSerMe	56
OY	241	CCTGGCTGGATGGCAATGCTGTTTCACGAGACAAGAATCCTTAGAGAGTTTTTGAAT	300

Alignment Scores:	
Pred. No.:	4.54e-05
Score:	129.00
Percent Similarity:	46.48%
Best Local Similarity:	23.84%
Query Match:	8.66%
DB:	5
Length:	12
Matches:	33
Conservative:	33
Mismatches:	32
Indels:	24
Gaps:	2

Db	85	Thv1alThnGlYThrLeuIleIleAspGlyGlnAlaLys-----AsnGlnLeuLys	101
QY	446	TTTCATCTGACACGCCCGAGGCTCCACCCAGCAACACAGTGTGGAAATGCCAGTACTGC	505
Db	102	Phv1aGlnValIpheAsnLeuAlaSerAsnAsnGlySerPheLeuIleAsnAspPhe	121
QY	506	TTTCCGC 511	
Db	122	PheArg 123	
RESULT 5			
	Q9P8H0	PRELIMINARY;	PRT; 123 AA.
ID	Q9P8H0		
AC	Q9P8H0		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Rpl1.		
GN	RPH1.		
OS	Yarrowia lipolytica (Candida lipolytica).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Dipodascaceae; Yarrowia.		
OX	NCBI_TaxID=4952;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Cordero Otero R.R., Leplingle A., Gallardin C.;		
RT	"RPH1 gene from Yarrowia lipolytica".		
RL	Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF260231; AAF70316.1; -		
DR	HSSP; P13662; IOUN.		
DR	InterPro; IPR002075; NTF2.		
DR	Pfam; PF02136; NTF2; 1.		
SO	SEQUENCE 123 AA; 13696 MW; 3E6CB95CF0058251 CRC64;		
Alignment Scores:			
	Pred. No.:	0.000108	Length: 123
	Score:	125.50	Matches: 40
	Percent Similarity:	44.78%	Conservative: 20
	Best Local Similarity:	29.85%	Mismatches: 57
	Query Match:	8.43%	Indels: 17
DB:		3	Gaps: 5
US-09-763-902B-10 (1-830) x Q9P8H0 (1-123)			
QY	116	TCGTGGATTTCAGAGCATATGTGATTCAGGCGCTGCAGACGTGCTGAGAGATTTCAT	175
Db	2	SerValAspPheAsnThrLeuAlaLysGlnPheCys-----GlnPhe-----	15
QY	176	GTTACTACACACACATGATAGGAGCGCGCTTTGCTGTCCCGCTTACATGAGGACA	235
Db	16	---TyrTyrGlnThrPheAspThrAspArgSerGlnLeuGlnLeuIleTyrArgAspHis	34
QY	236	GCCACCCCTGCTCGAATGAGCAATGCTGTTTCAGAGCAAGAATCCTTGAGTAAATTTT	295
		:::     :::	:::     :::
Db	35	SerMetLeuThrPheThrGlyThrGlnHisGlnGlnAlaIleValGlnLysLeu	54
QY	296	GAATGTGCTCCCTCCAGCAG-----TTCCAATCAGCGTGTGAGTACTGCCACCTGTT	349
Db	55	ValGlyLeuProPheGlyGlnValArgHisIstysIleSerAspIleAspAlaGlnPro---	73
QY	350	CATGATGATGACACACACAGCCAGACACACAGCGTCTGTGTGCATCTGTGATCAGTGA	409
Db	74	-----AlaSerIleGlnGlyAspValIleValIleValThrGlyLysLeuCys	90
QY	410	TTTGAGGGGAAACAACAGCGGACTTCACACAGATCATCTATCATGACCGCCAGCGCTCA	469
		:::     :::	:::     :::
Db	91	ValAspGlyAspAsnProLeuProTyrGlyGlnValPheHisIleLeu-----	106
QY	470	CCGAGCAACACAGTGTGAAGATCGCAACTGACTCTCCGC 511	



US-09-763-902B-10 (1-830) x Q9NH72 (1-690)

QY 164 GAATTTGTCAATGTCTACTACACCAACATGATTAAGCGGGCGGCTTTCTCTCCGCGCTG 22

Db 19 GlnPheValaTargInTyrTyrLeuLeuAsnLysAlaProAsnHisLeuHisArgPhe 38

QY 224 TAAATGGGCAAGCCACCCCTGGTCC--TGGAATGGCAATGCGTTTACGACAAAGATTC 280

Db 39 TTAAsnHisLeuAsnSerSerTyrTyrLeuHisTslGlyuSerLysLeuValaValGlyGlnArgGlu 58

QY 281 TTAGTGTGAGTTTTTTTAAATGTTCCTTCCTCCACCGAG-----TTCCAAATCGCGTGTA 334

Db 59 ILSHISAsnArgTleInclInclInLeuAsnPhenAsnAryCSHISAlaLysIleSerGlnVal 78

QY 335 GACTGGCCAGCCTGTTCATGATGATGAAGCCACACACACAGCCAGACCGCTCTGTGTGATC 394

Db 79 AspAlaGlnAlaThrLeuGlnLysAsnGly-----ValaValaGlnVal 92

QY 395 TGTGGATCACTGAAGTTTATGAGGGGAACAACAAGCGAGCTTCACCAACTTCATCTG 45

Db 93 ThnGlyGlnLeuSerAsnAspGlyGlnPrometaArgArgPheThrGlnThrPheValLeu 112

QY 455 ACCGCCAGGCGCCACCCACCAACAACAGTGTGATGATGCGCAAGTGCCTCGGCTG 514

Db 113 AlalAlaGln-----SerProLysLys-----TyrTyrValHisAsnAspIlePheArgTyr 12

QY 515 CAGGAC 520

Db 130 Glnasp 131

RESULT 8

Q9VFT4 PRELIMINARY: PRT: 690 AA.

Q9VFT4

AC Q9VFT4: 01-MAY-2000 (TREMBLrel, 13, Created)

DT 01-MAY-2000 (TREMBLrel, 13, Last sequence update)

DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)

DE CG9412 protein (L031194P).

GN RIN OR CG9412.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

NCBI\_faxid=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY.

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballen R.M., Basu A.A., Baxendale J., Bayraktaroglu L., Beaskey E.M.,

RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,

RA Bokorva D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,

RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Dublin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Flockerzi C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Gloder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Idbewam C.,

RA Jatalil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Mekulov G., Mishina N.V., Moberly C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA palazzolo M., Pittman G.S., Pan S., Pollard J., Put V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Slieden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.X.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nunoo J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,  
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE003701; AAF54955.1;  
 DR EMBL: AY058617; AAL13845.1;  
 DR FlyBase: FBgn0015778; rin.  
 DR InterPro: IPR002075; NTF2.  
 DR InterPro: IPR005054; RNA\_rec\_mot.  
 DR Pfam: PF02136; NTF2; 1.  
 DR Pfam: PF00076; rim; 1.  
 DR SMART: SMO0360; RRM; 1.  
 DR PROSITE: PSS0102; RRM; 1.  
 SQ SEQUENCE 690 AA; 74943 MW; 40C5AC0F9FAEB53 CRC64;  
  
 Alignment Scores:  
 Pred. No.: 0.00803 Length: 690  
 Score: 119.00 Matches: 37  
 Percent Similarity: 45.90% Conservative: 19  
 Best Local Similarity: 30.33% Mismatches: 54  
 Query Match: 7.99% Indels: 12  
 DB: 5 Gaps: 5  
  
 US-09-763-902B-10 (1-830) x Q9VFT4 (1-690)  
 QY 164 GAGTTTGTCATAGTCTACTATACACACCATGATTAAGCGCGCGTTCCTGCCCGCTG 223  
 Db 19 GIUPheValaIargGIIntYtrTyrThrIleuAsnIlyAlaPRoAsnHIsIeuHIsArgPhe 38  
 QY 224 TACATGGGCACACGCCACCTGCTC---TGGAAATGGCAATGCTTTTCAGACACAAGATCC 280  
 Db 39 TYraSnHIsAsnIsSerSerTyrTrIleHIsIGlyIuSerIyLysIeuValaIaIGlyGlnaIrgIu 58  
 QY 281 TTGAGTGAAGTTTTCGAATGTTGCTTCAGCTCCAGCGAG-----TTCCAATCAGCGTGTA 334  
 Db 59 ILeHIsAsnArgIleGIInIleuAsnPRheAsnAsPCySHIsaIaIySIIeSeGIInVal 78  
 QY 335 GACTGCCACGCCCTTCATCATGATGAGCCACACAGCCAGACGCCAGCGCTCTGTGTGCATC 394  
 Db 79 AsPaIaIGlnaIaIThrIleuIyLysnGly-----ValValaIaIGlnaIVal 92  
 QY 395 TGTGATCATGTCAGATTGAGGGGAACAACAACGGAGCTTCAACAGAACTTCATCTG 454  
 Db 93 ThrgIyGIuIeSerAsnAspIyGlnPRomeTArgARpHeThrgIInThrPRheValIeu 112  
 QY 455 ACCGGCAGCGCTCAGCCAGACAACACAGTGTGAATCCAGTACGCTGCTCCGCTG 514  
 Db 113 AlaIaIaGln---SerProIyLys-----TyrTyrValHIsAsnAspIlePRheIyGlyr 129  
 QY 515 CAGGAC 520  
 Db 130 GlnAsp 131

ID 09LPY0 PRELIMINARY; PRT; 522 AA.  
AC 09LPY0: PRELIMINARY; PRT; 522 AA.  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE T23J18.22.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shum P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,  
RA Kim C., Alatai H., Bei Q., Chin C., Chion J., Choi E., Conn L.,  
RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharzky N., Nguyen M.,  
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thayerl A.,  
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
RA Ecker J.R.;  
RT "Genomic sequence for Arabidopsis thaliana BAC T23J18 from chromosome  
RT 1."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AC011661; AAF16635.1;  
DR HSSP: P13662; 10UN.  
DR InterPro: IPR002075; NTF2.  
DR Pfam: PF02136; NTF2; 1.  
SQ SEQUENCE 522 AA; 58484 MW; 74FB3FDEE5B86 CRC64;  
Alignment Scores:  
Pred. No.: 0.000968 Length: 522  
Score: 118.00 Matches: 38  
Percent Similarity: 43.44% Conservative: 15  
Best Local Similarity: 31.15% Mismatches: 49  
Query Match: 7.92% Indels: 20  
DB: 10 Gaps: 4  
US-09-763-902B-10 (1-830) x 09LPY0 (1-522)  
QY 98 ATAGAACGAGATGGCATCTGTGATTTCAAGACCTATGTGATCAGCGCTGACAGCT 157  
DB 335 IIEGLUGLUGLVALALASERALA----- 342  
QY 158 GCTGAGAGATTGTCAATGCTCTACACACACCATGATACGGCGCGGTTGCTGCC 217  
DB 343 -----PHEVALASINHSITRYTHLSLEUPHEASPNASPARSERLEUSER 359  
QY 218 CGCCTGATCATGGGACACACCGCTGCTGATGGCATGCTGTTTCAGACAGAA 277  
DB 360 SerLeuTYAsnProThrSerLeuLeuThrPheGLUGLGLNThrILeTYGLYAlaAsp 379  
QY 278 TCCTTGAGTGAATTTTGAATGTTGCCTTCACAGAGTTCCAA-----ATCAGCGTG 331  
DB 380 AsnIleSerAsnLYsLeuLYsGLNLeuProPheAspGLNcYSHISLeuIleSerThr 399  
QY 332 GTAGACTGCCACCGCTTTCATATGAGACACACACACACACACGCTGCTGTTGTC 391  
DB 400 ValAspSerGLNPro-----SerSerMetAlaGLYGLYcYsGLYLeuValPhe 417  
QY 392 ATCTGTGATCAGTGAAGTTGAGGGG---AACAAACACAGGAGCTTCAACAGAACTTC 448  
DB 418 ValSerGLYSerIleGLNLeuHISGLYAlaAspHISProLeuAlaRPhSerSerGLNThrPhe 437  
QY 449 ATCTGTG 454  
DB 438 HistLeu 439  
RESULT 10  
ID 096VN3 PRELIMINARY; PRT; 125 AA.  
AC 096VN3: PRELIMINARY; PRT; 125 AA.  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Nuclear transport factor 2.  
CN NTF2.  
OS Emericella nidulans (Aspergillus nidulans).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; Emericella.  
OX NCBI\_TaxID=5072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Espeso E.A., Penalva M.A.;  
RT "Nuclear transport factor 2 from Aspergillus nidulans."  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AY038983; AAK71467.1;  
DR InterPro: IPR002075; NTF2.  
DR Pfam: PF02136; NTF2; 1.  
SQ SEQUENCE 125 AA; 13871 MW; 8FC18BDFC597E2F2 CRC64;  
Alignment Scores:  
Pred. No.: 0.00278 Length: 125  
Score: 112.50 Matches: 33  
Percent Similarity: 47.11% Conservative: 24  
Best Local Similarity: 27.27% Mismatches: 53  
Query Match: 7.56% Indels: 11  
DB: 3 Gaps: 4  
US-09-763-902B-10 (1-830) x 096VN3 (1-125)  
QY 158 GCTGAGAGATTGTCAATGCTCTACACACACCATGATACGGCGCGGTTGCTGCC 217  
DB 8 AlAGLNGlNpHeValThrPheTYrTYrGLNThrPheAspGLYAsnArgAlaGLYLeuAla 27  
QY 218 CGCCTGATCATGGGACACACCGCTGCTGATGGCATGCTGTTTCAGACAGAA 277  
DB 28 ProLeuTYArgAspHISerMetLeuThrPheGLNThrSerAlaIleGLNLYAla 47  
QY 278 TCCTTGAGTGAATTTTGAATGTTGCCT-----TCAGGAGATGCCAATACGCGTG 331  
DB 48 GLYIleIleGLYsLeuThrSerLeuProPheGLNLYsValGLNHSISLInValSerThr 67  
QY 332 GTAGACTGCCACCGCTTTCATATGATGATACACACACACACACGCTGCTGTTGTC 391  
DB 68 LeuAspAlaGLNProSerSerGLYcYsGLY-----GLYLeuValLeu 82  
QY 392 ATCTGTGATCAGTGAAGTTGAGGGGGAACAACA---CGGAGCTTCAACAGAACTTC 448  
DB 83 ValThrGLYAlaLeuLeuValAspGLNGLYsAsnProMetAlaTYrThrGLNThrPhe 102  
QY 449 ATCTGTGACCGCCACAGCGCTCAGCCACACACACAGTGTGGAAGATGCCAGTGTCTTC 508  
DB 103 GLNLeuMetProAspGLYAlaGLYSer-----TYrPheValLeuAsnAspValPhe 119  
QY 509 CGC 511  
DB 120 Arg 120  
RESULT 11  
ID 021351 PRELIMINARY; PRT; 537 AA.  
AC 021351: PRELIMINARY; PRT; 537 AA.  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE K08F4.2 protein.  
GN K08F4.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Felodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hemry C.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]

RA George R.A., Lewis S.E., Richards S.J., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H.C., Blazaj R.G., Champagne K., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abiril J.F., Adayanti A., An H.-J., Andrews-Plamkoch C., Baldwin D.,  
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dopp L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Garbrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
RA Jalali M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostreil A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacible J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Splitter E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
RA Svrtikas R., Teclor C., Turner R., Venler E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000)  
DR EMBL; AF003659; AAP53669.1; -.  
DR HSSP; P13662; 10UN.  
DR FLybase; FBgn0032680; CG10174.  
DR InterPro; IPRO02075; NTF2.  
DR Pfam; PF02136; NTF2; 1.  
SQ SEQUENCE 130 AA; 14582 MW; 39286BFC5BD7C34E CRC64;

Alignment Scores:  
Pred. No.: 0.0161 Length: 130  
Score: 105.50 Matches: 35  
Percent Similarity: 43.20% Conservative: 19  
Best Local Similarity: 28.00% Mismatches: 50  
Query Match: 7.09% Indels: 21  
Gaps: 6

US-09-763-902B-10 (1-830) x Q9VJ85 (1-130)

QY 161 GAGGAGTGTCTCAATGTCTACTACACCACCATGGAT-----AAGCGGCGGCGTTTGCTG 214  
DB ::::||||| 1111111 ||||| 1111111 :::  
12 LysGIuIPheValGIInIrrYrAlaIIePheAsPaSPPrOlaIasAArGuIAuSvAl 31  
QY 215 TCCCGCCTTCATCATGGGCACA---GCCACCGCTGCTGGAAGTGAAGTGCCTGTTACAGA 274  
DB :||| ||| ::::: ||| :||| |::: |||  
32 IleasnPheryrAsnAlaIhrspSePIhemerThrPheoIuGlYnsngInIIegIngly 51  
QY 272 CAAGATCCCTGTAGTAGTATTTTGTGAATGTCCTCCCTCCACGAGATCTCAA----- 322  
DB :::: ||| :||| :||| :||| :||| :|||  
52 AlaPolYsIlle-----LeuGIuIlysvAlGIInserrIeuSerPheGInIlyIleAla 68  
QY 323 -----ATCAGCGGTGTAGACTGCCAGCGCTGTTCATGATGTAAGCACCAAGCCAGACCC 376  
DB :|||::: ||||| ||||| ||||| |||  
69 ArgValIIeThrValaspserGIInPrOthrSerAspIlylly----- 83  
QY 377 AGCGTCCTGTGTCTATCTGTGATATGATGTAAGTATTGAGGGGAACAACAAGCGGACTTC 438  
DB :|||::: ||||| ||||| ||||| |||  
84 ---ValIeuIIeIleValIeugIlyArGleuLyIsCyAsPaSPspPrOPrOHISAlaIphe 102

QY 437 AACGAGACTTCCTGACCCCGCCAGCGCTCACCCACACACAGTGTGGAGATCGCA 496  
 Db 103 SerGlnIlePheLeuLeu-----LysProAsnGlySerLeuPheValala 118  
 QY 497 AGTGAAGTCTCCCGC 511  
 Db 119 HisAspIlePheArg 123

RESULT 13  
 Q9FZK4  
 ID Q9FZK4 PRELIMINARY: PRT; 122 AA.  
 AC Q9FZK4  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE F17L21.10 (Similar to nuclear transport factor 2).  
 GN F17L21.10.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Johnson-Hopson C., Brooks S., Buehler E., Chao Q., Khan S., Kim C.,  
 RA Shinn P., Altafi H., Bei B., Chin C., Chlou J., Choi E., Conn L.,  
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.;  
 RT "Genomic sequence for Arabidopsis thaliana BAC F17L21 from chromosome  
 RT I.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chlou J., Choi E.,  
 RA Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,

RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL: AC004557; AAF99749.1;  
 DR EMBL: AY042889; AAK68829.1;  
 DR EMBL: AY072473; AAL68888.1;  
 DR HSP: F13662; IOUN.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2; 1.  
 SQ SEQUENCE 122 AA: 13527 MW; E7CDD9486631A1D2 CRC64;

Alignment Scores:  
 Pred. No.: 0.018 Length: 122  
 Score: 105.00 Matches: 33  
 Percent Similarity: 45.45% Conservative: 22  
 Best Local Similarity: 27.27% Mismatches: 54  
 Query Match: 7.05% Indels: 12  
 DB: 10 Gaps: 4

US-09-763-902B-10 (1-830) x Q9FZK4 (1-122)

QY 158 GCTGAGGAGTTGTCTCAATGTCTACTACACCACCATGGATACGGCGGCGTTGCTGCTC 217  
 Db 7 AlalysAlaPheValGluHisTyrSerThrPheAspAlaAsnArgProGlyLeuVal 26  
 QY 218 CGCTGTACATGGCGACACGCCACCTGGTCTCGAATGGCAATGCTTTTCAGGACAGAA 277  
 Db 27 SerLeuTyrGlnGluGlySerMetLeuThrPheGluGlyGlnGlySerGln 46  
 QY 278 TCCTTGAGTGAGTGTGTTTGAATGTGCTCT-----TCCAGCGAGTTCACAAATCAGCGTG 331  
 Db 47 AsnIleValAlaLysLeuThrGlyLeuProPheGlnGlnCysLysHisAsnIleThr 66  
 QY 332 GTAGACTGCCAGCCCTGTCATGATGAAGCCACACCAAGCCAGCCAGCTCTGTTGTC 391  
 Db 67 ValAspCysGlnPro-----SerGlyProAlaGlyGlyMetLeuValPhe 81  
 QY 392 ATCTGTGATCAGTGAAGTTTCAGGGGACACAAACACGG--GACTTCAACCAAGACTTC 448  
 Db 82 ValSerGlyAsnLeuGlnLeuAlaGlyGluGlnHisAlaLeuLysPheSerGlnMetPhe 101  
 QY 449 ATCTGACCGCCCGCCCTCACCAGACACAGTGTGGAGATCGCAAGTGAAGTCTTC 508  
 Db 102 HisLeuIleSerAsnGlnGlyAsn-----TyrTyrValPheAsnAspIlePhe 117  
 QY 509 CGC 511  
 Db 118 Arg 118

RESULT 14  
 Q9HFS7  
 ID Q9HFS7 PRELIMINARY: PRT; 796 AA.  
 AC Q9HFS7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Deubiquitinating enzyme.  
 OS Kluyveromyces lactis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Amerik A., Li S.-J., Hochstrasser M.;  
 RT "Analysis of the deubiquitinating enzymes of the yeast Saccharomyces  
 RT cerevisiae";  
 RL Biol. Chem. 0:0-0(2000).  
 DR EMBL: AF303215; RAG17929.1;  
 DR InterPro: IPR001763; Rhodanese-like.  
 DR InterPro: IPR001394; UCH-2.  
 DR Pfam: PF00442; UCH-1; 1.





Mon Jun 30 08:51:03.2003

us-09-763-902b-10.n2p.rsp

Page 10

[illegible]

Search completed: June 24, 2003, 19:36:51  
Job time : 64.125 secs

1	978	85.3	1387	10	US-09-766-864-393	Sequence 393, App
2	747.2	65.2	8918	10	US-09-764-884-1682	Sequence 1682, App
3	747.2	65.2	8919	10	US-09-764-864-1683	Sequence 1683, App
4	414.8	36.2	464	9	US-09-796-659-8289	Sequence 8289, App
5	414.8	36.2	464	9	US-10-040-862-8289	Sequence 8289, App
6	251.6	22.0	864	10	US-09-764-864-394	Sequence 394, App
7	184.8	16.1	318	10	US-09-764-343-1201	Sequence 1201, App
8	184.2	16.1	318	9	US-09-796-692-4160	Sequence 4160, App
9	184.2	16.1	318	9	US-10-040-862-4160	Sequence 4160, App
10	86.2	7.5	415	9	US-09-918-999-2901	Sequence 2901, App
11	62	3.4	417	9	US-09-918-999-6557	Sequence 6557, App
12	43.8	3.8	15732	9	US-10-239-676-96	Sequence 96, App
13	43	3.8	5689	9	US-10-239-676-90	Sequence 90, App
14	42.8	3.7	11812	9	US-10-239-676-209	Sequence 209, App
15	42.8	3.7	18997	9	US-10-172-068-18	Sequence 18, App
16	41.6	3.6	6018	9	US-10-239-676-92	Sequence 92, App
17	41.6	3.6	8842	9	US-10-239-676-72	Sequence 72, App
18	41	3.6	7195	9	US-10-239-676-30	Sequence 30, App
19	40.8	3.6	17848	9	US-10-239-676-27	Sequence 27, App

C	20	40.4	3.5	640681	10	US-09-790-888-1	Sequence 1, App1
C	21	40.2	3.5	9254	9	US-10-239-676-85	Sequence 85, App1
C	22	40	3.5	254366	9	US-09-822-871-3	Sequence 3, App1
C	23	39.2	3.4	2716	10	US-09-764-664-191	Sequence 191, App1
C	24	39	3.4	5951	9	US-10-239-676-115	Sequence 115, App
C	25	38.8	3.4	4747	9	US-09-764-891-6305	Sequence 6305, App
C	26	38.8	3.4	4747	9	US-10-205-428-610	Sequence 610, App
C	27	38.6	3.4	399	10	US-09-960-352-10222	Sequence 10222, App
C	28	38.4	3.4	17421	9	US-10-239-676-54	Sequence 54, App1
C	29	38.4	3.4	26006	9	US-10-091-504-1963	Sequence 1963, App
C	30	38.4	3.4	26006	9	US-09-764-891-6585	Sequence 6585, App
C	31	38.4	3.4	26006	10	US-09-764-869-1963	Sequence 1963, App
C	32	38.4	3.4	26006	10	US-09-764-864-1638	Sequence 1638, App
C	33	38.4	3.4	26013	9	US-10-091-504-1961	Sequence 1961, App
C	34	38.4	3.4	26013	9	US-09-764-891-6283	Sequence 6283, App
C	35	38.4	3.4	26013	10	US-09-764-869-1961	Sequence 1961, App
C	36	38.4	3.4	26013	10	US-09-764-864-1636	Sequence 1636, App
C	37	38.4	3.4	26018	9	US-10-091-504-1962	Sequence 1962, App
C	38	38.4	3.4	26018	9	US-09-764-891-6284	Sequence 6284, App
C	39	38.4	3.4	26018	10	US-09-764-869-1962	Sequence 1962, App
C	40	38.4	3.4	26018	10	US-09-764-864-1637	Sequence 1637, App
C	41	38.2	3.3	12194	9	US-10-091-438-283	Sequence 283, App
C	42	38.2	3.3	17419	9	US-10-239-676-99	Sequence 99, App1
C	43	38	3.3	11812	9	US-10-239-676-210	Sequence 210, App
C	44	37.8	3.3	175	10	US-09-969-373-50	Sequence 50, App
C	45	37.6	3.3	671	9	US-10-184-644-346	Sequence 346, App

## ALIGNMENTS

RESULT :  
HS-09-70

```

Sequence 393, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 393
LENGTH: 1387
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1379)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (1380)
OTHER INFORMATION: n equals a,t,g, or c
IS-09-764-864-393

```

	Query Match	85.3%	Score 978;	DB 10;	Length 1387;
	Best Local Similarity	100.0%;	Pred. No. 1.7e-240;		
	Matches 978;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY	169	GGATTTTAAACTTATGTAGATCAGCATGTAGACCTGCTGAGCAGTTTGTCAATATTTA	222		
Db	75	GGATTTTAAACTTATGTAGATCAGCATGTAGACCTGCTGAGCAGTTTGTCAATATTTA	134		
QY	229	CTATGACGAATGCTATTAAGACGGGCACTAACCGCTGTATCTGACACAGGCCAC	288		
Db	135	CTATGACGAATGCTATTAAGACGGGCACTAACCGCTGTATCTGACACAGGCCAC	194		
QY	289	CTTATATGGAATGGAATGCTGTTTCAGGGCTGGATGCCCTTAATATATTTTGTGACAC	348		
Db	135	CTTATATGGAATGGAATGCTGTTTCAGGGCTGGATGCCCTTAATATATTTTGTGACAC	254		

```

OY 349 ATGCTCTAGTGTGAGTCCAGGTCATATGTATGTCACCAACAGTTCATGACGAC 408
    |||||
DB 255 ATGGCTCTAGTGTGAGTCCAGGTCATATGTATGTCACCAACAGTTCATGACGAC 314
OY 409 AACTCAGTCCCAACTACAGTCTGTGTGTGACAGGTCGAACTGGAAGTTGATGAAA 468
    |||||
DB 315 AACTCAGTCCCAACTACAGTCTGTGTGTGACAGGTCGAACTGGAAGTTGATGAAA 374
OY 469 CAACAACTTTCTTCAACAGCAACTCTCTGTGTGACAGTTCATGACGAC 528
    |||||
DB 375 CAACAACTTTCTTCAACAGCAACTCTCTGTGTGACAGTTCATGACGAC 434
OY 529 TGTGTGAGAGATGCAAGTGTGTCCTGCTTTCAGGTCGTCATGATGTAAGG 588
    |||||
DB 435 TGTGTGAGAGATGCAAGTGTGTCCTGCTTTCAGGTCGTCATGATGTAAGG 494
OY 589 CAACAGTCCATTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 648
    |||||
DB 495 CAACAGTCCATTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 554
OY 649 TGAATGAGAGAGTATATATATGTCGAACTAAATTTCTTAAATTTCTTAA 708
    |||||
DB 555 TGAATGAGAGAGTATATATATGTCGAACTAAATTTCTTAAATTTCTTAA 614
OY 709 GTGACAGCTTTCTGACAGCTGCGAGTGTGAGCATGCGCTTAAGGCTTAA 768
    |||||
DB 615 GTGACAGCTTTCTGACAGCTGCGAGTGTGAGCATGCGCTTAAGGCTTAA 674
OY 769 ATTTTTCATGCTTATATATATGTCATGTCATGTCATGTCATGTCATGTC 828
    |||||
DB 675 ATTTTTCATGCTTATATATATGTCATGTCATGTCATGTCATGTCATGTC 734
OY 829 ATGACCTTGTCTAATATATGTCATGTCATGTCATGTCATGTCATGTCAT 888
    |||||
DB 735 ATGACCTTGTCTAATATATGTCATGTCATGTCATGTCATGTCATGTCAT 794
OY 889 TATATATATATATATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 948
    |||||
DB 795 TATATATATATATATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 854
OY 949 TGGTATATAGTAAATTTCTGATGAAACAGCTGCTGCTTTCAGTTTAA 1008
    |||||
DB 855 TGGTATATAGTAAATTTCTGATGAAACAGCTGCTGCTTTCAGTTTAA 914
OY 1009 GCCATTTTATATATGCTTTTCAATATGTCATGTCATGTCATGTCATG 1068
    |||||
DB 915 GCCATTTTATATATGCTTTTCAATATGTCATGTCATGTCATGTCATG 974
OY 1069 ACATATCTCTGATATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1128
    |||||
DB 975 ACATATCTCTGATATGTCATGTCATGTCATGTCATGTCATGTCATGTC 1034
OY 1129 TTAAGTAAATTTTGTGTTT 1146
    |||||
DB 1035 TTAAGTAAATTTTGTGTTT 1052
    |||||

RESULT 2
US-09-764-864-1682/c
Sequence 1682, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1682
LENGTH: 8918
TYPE: DNA

```

```

; ORGANISM: Homo sapiens
US-09-764-864-1682
Query Match 65.2% Score 747.2; DB 10; Length 8918;
Best Local Similarity 99.6%; Fred. No. 7e-181;
Matches 749; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 395 GTTCATGACAGCAAGTCTGATGTCATGTCATGTCATGTCATGTCATGTC 454
    |||||
DB 2199 GATGAGAGCAAGCAAGTCTGATGTCATGTCATGTCATGTCATGTCATGTC 2140
OY 455 AAGTTGATGGAACCAACCAATTTCTTCAACAGCAAGTCTGATGTCATGTC 514
    |||||
DB 2139 AAGTTGATGGAACCAACCAATTTCTTCAACAGCAAGTCTGATGTCATGTC 2080
OY 515 ACTCCCAACAACTACTGTGTGAGAGTTCGAAGTGTGTCCTGCTTTCAGAT 574
    |||||
DB 2079 ACTCCCAACAACTACTGTGTGAGAGTTCGAAGTGTGTCCTGCTTTCAGAT 2020
OY 575 AGTATTAAGGGGCAAGAGTCTGATGTCATGTCATGTCATGTCATGTCATG 634
    |||||
DB 2019 AGTATTAAGGGGCAAGAGTCTGATGTCATGTCATGTCATGTCATGTCAT 1960
OY 635 TATGTAATTTATTTGATTTGATGAGCACTATATATGTCGTAAGTAAAT 694
    |||||
DB 1959 TATGTAATTTATTTGATTTGATGAGCACTATATATGTCGTAAGTAAAT 1900
OY 695 TATTTTCATTCCTGTCAGACACTTTTCTAGACGTCGACGTCGTCGTCGTC 754
    |||||
DB 1899 TATTTTCATTCCTGTCAGACACTTTTCTAGACGTCGACGTCGTCGTCGTC 1840
OY 755 AGAGCTTAAACCTATTTTTCATGTCATGTCATGTCATGTCATGTCATGTC 814
    |||||
DB 1839 AGAGCTTAAACCTATTTTTCATGTCATGTCATGTCATGTCATGTCATGTC 1780
OY 815 ATATATATTAACACATGATCTGTGTAACATATGTCATGTCATGTCATGTC 874
    |||||
DB 1779 ATATATATTAACACATGATCTGTGTAACATATGTCATGTCATGTCATGTC 1720
OY 875 AATATATCTTTTATATATATATATATGTCATGTCATGTCATGTCATGTC 934
    |||||
DB 1719 AATATATCTTTTATATATATATATATGTCATGTCATGTCATGTCATGTC 1660
OY 935 GAAGTGAAGTTTGTGTTATGTTATGTTATGTTATGTTATGTTATGTTATG 994
    |||||
DB 1659 GAAGTGAAGTTTGTGTTATGTTATGTTATGTTATGTTATGTTATGTTATG 1600
OY 995 TTAACACTGTATATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 1054
    |||||
DB 1599 TTAACACTGTATATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 1540
OY 1055 ACAGATTTAGGTTTACATATGTCATGTCATGTCATGTCATGTCATGTCATG 1114
    |||||
DB 1539 ACAGATTTAGGTTTACATATGTCATGTCATGTCATGTCATGTCATGTCATG 1480
OY 1115 GGTGCTTTTCAATTTTAAATTTTGTGTTT 1146
    |||||
DB 1479 GGTGCTTTTCAATTTTAAATTTTGTGTTT 1448
    |||||

RESULT 3
US-09-764-864-1683/c
Sequence 1683, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 1683
; LENGTH: 8919
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-1683

```

Query Match	65.2%	Score 747.2;	DB 10;	Length 8919;
Best Local Similarity	99.6%	Pred. No. 7e-181;		
Matches 749; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

QY	395	GTTCATATGACGACAACTACACCTCCAAACTACAGTCTTGTTGACACAGTGAACCTG	454
Db	2200	GATGAAGGACAGCACTCAGTCCCAACTACAGTCTTGTTGACACAGTGAACCTG	214
QY	455	AAGTTGTATGGAACCAACCAACATTTCTTCAACAGAACTTCTGCTGACTGCTCAGTCC	514
Db	2140	AAGTTGTATGGAACCAACCAACATTTCTTCAACAGAACTTCTGCTGACTGCTCAGTCC	208
QY	515	ACTCCCAACATTA CTGTGTGGAGAATTCGAAGTATGTGCTTCGGTTTTCAGATTTGCT	574
Db	2080	ACTCCCAACATTA CTGTGTGGAGAATTCGAAGTATGTGCTTCGGTTTTCAGATTTGCT	202
QY	575	AGTAGTTAAAGGGGCAAAAGTCATTCATTTGGTTCATTTAGTTTCAGACATTTGAAT	634
Db	2020	AGTAGTTAAAGGGGCAAAAGTCATTCATTTGGTTCATTTAGTTTCAGACATTTGAAT	196
QY	635	TATGTGAATATTTTGTATGTATGTAAGAACAATATATATGTGCTGAACATAATTTCTTAA	694
Db	1960	TATGTGAATATTTTGTATGTATGTAAGAACAATATATATGTGCTGAACATAATTTCTTAA	190
QY	695	TATTTTCATTCCTGTCAGACACTTTTCTAGAGCTGCGACAGTTTGAGAGATTCGCCCTCA	754
Db	1900	TATTTTCATTCCTGTCAGACACTTTTCTAGAGCTGCGACAGTTTGAGAGATTCGCCCTCA	184
QY	755	AGAGCTTAAACATTTTTTTACATGCGCTTATATACATTCACATTAATGACATTCCTAA	814
Db	1840	AGAGCTTAAACATTTTTTTACATGCGCTTATATACATTCACATTAATGACATTCCTAA	178
QY	815	ATAATATTAAACACATGATCTTGGTACTAACAATACTACTGTAACCCAGCCTATTGCAA	874
Db	1780	ATAATATTAAACACATGATCTTGGTACTAACAATACTACTGTAACCCAGCCTATTGCAA	172
QY	875	AAATPAAATCTTTTATATATATTATCTATGAGATGTCAGACAAATPATAACTCTGGGAA	934
Db	1720	AAATPAAATCTTTTATATATATTATCTATGAGATGTCAGACAAATPATAACTCTGGGAA	166
QY	935	GAAGTGGAGTTTTTGGTTATTAAGTTAAATTTTCTAGTAAACAGAGTTCGCTGTTTCAG	994
Db	1660	GAAGTGGAGTTTTTGGTTATTAAGTTAAATTTTCTAGTAAACAGAGTTCGCTGTTTCAG	160
QY	995	TTAACACGTAATGCCATTTTAAATATATGCTTTTCCAAATCAGAGTTCAGTGAATAAT	105
Db	1600	TTAACACGTAATGCCATTTTAAATATATGCTTTTCCAAATCAGAGTTCAGTGAATAAT	154
QY	1055	ACAGATTAGGTTTACATTA CTGTCGACATACAGTGAATTCATATAGAGATGTTCACT	111
Db	1540	ACAGATTAGGTTTACATTA CTGTCGACATACAGTGAATTCATATAGAGATGTTCACT	148
QY	1115	GGTCGTTTTCAATTTAAGTATTTTGTGTTT	1146
Db	1480	GGTCGTTTTCAATTTAAGTATTTTGTGTTT	1449

RESULT 4  
US-09-796-692-8289  
; Sequence 8289, Application US/09796692  
; Publication NO. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Galger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES

```

1 FILE REFERENCE: 2077.001200
2 CURRENT APPLICATION NUMBER: US/09/796,692
3 CURRENT FILING DATE: 2001-03-01
4 PRIOR APPLICATION NUMBER: 60/186,126
5 PRIOR FILING DATE: 2000-03-01
6 PRIOR APPLICATION NUMBER: 60/190,479
7 PRIOR FILING DATE: 2000-03-17
8 PRIOR APPLICATION NUMBER: 60/200,545
9 PRIOR FILING DATE: 2000-04-27
10 PRIOR APPLICATION NUMBER: 60/200,303
11 PRIOR FILING DATE: 2000-04-28
12 PRIOR APPLICATION NUMBER: 60/200,779
13 PRIOR FILING DATE: 2000-04-28
14 PRIOR APPLICATION NUMBER: 60/200,999
15 PRIOR FILING DATE: 2000-05-01
16 PRIOR APPLICATION NUMBER: 60/202,084
17 PRIOR FILING DATE: 2000-05-04
18 PRIOR APPLICATION NUMBER: 60/206,201
19 PRIOR FILING DATE: 2000-05-22
20 PRIOR APPLICATION NUMBER: 60/218,950
21 PRIOR FILING DATE: 2000-07-14
22 PRIOR APPLICATION NUMBER: 60/222,903
23 PRIOR FILING DATE: 2000-08-03
24 PRIOR APPLICATION NUMBER: 60/223,416
25 PRIOR FILING DATE: 2000-08-04
26 PRIOR APPLICATION NUMBER: 60/223,378
27 PRIOR FILING DATE: 2000-08-07
28 NUMBER OF SEQ ID NOS: 9597
29 SOFTWARE: fastseq for windows Version 3.0
30 SEQ ID NO 8289
31 LENGTH: 464
32 TYPE: DNA
33 ORGANISM: Homo sapiens
34 US-09-796-692-8289

```

Query Match	36.2%	Score	414.8	DB	9	Length	466
Best Local Similarity	99.5%	Pred. No.	2.1e-96				
Matches	416	Conservative	2	Mismatches	0	Indels	0
						Gaps	0

QY 729 CTGCCAGTTTGAGAGATTGCCCTCTAAGAGCTTTAAACTATTTTTTTCAGTGCCTTAA 788  
 Db 1 CTGCCAGTTTGAGAGATTGCCCTCTAAGAGCTTTAAACTATTTTTTTCAGTGCCTTAA 60  
 QY 789 TACATTCACATATATACATTTCTTAATAATATTTAAACACATGATCTTGCTACTAACA 848  
 Db 61 TACATTCACATATATATACATTTCTTAATAATATTTAAACACATGATCCTTGCTACTAACA 120  
 QY 849 CTCACCTGTGAACCCAGCCCTATTGCACAAAATAAATCTTTTATTAATATTATCTATGGGAT 908  
 Db 121 CTCACCTGTGAACCCAGCCCTATTGCACAAAATAAATCTTTTATTAATATTATCTATGGGAT 180  
 QY 909 GTACACACAAATATAACACTCTCTGGGAAGAAGTGAGAGTTTTTGCTTATTTAGGTTAATTTTC 968  
 Db 181 GTACACACAAATATAACACTCTCTGGGAAGAAGTGAGAGTTTTTGCTTATTTAGGTTAATTTTC 240  
 QY 969 TAGTAAACACGCTTCGCTGTTTTCAGTTAAACACTGCTAATGTCATTTAATATATGCTT 1021  
 Db 241 TAGTAAACACGCTTCGCTGTTTTCAGTTAAACACTGCTAATGTCATTTAATATATGCTT 300  
 QY 1029 TTTCAAATCAGTTCTAGTGAAAATAGTATACAGATTATAGTTTACATAACTACTCTGCATATAC 1081  
 Db 301 TTTCAAATCAGTTCTAGTGAAAATAGTATACAGATTATAGTTTACATAACTACTCTGCATATAC 360  
 QY 1089 TGAATATGCATATAGAGATGTTCAGGTGCTGCTTTTTCATATTTAAGTAATTTTTTGTTTT 1146  
 Db 361 TGAATATGCATATAGAGATGTTCAGGTGCTGCTTTTTCATATTTAAGTAATTTTTTGTTTT 418

```

RESULT5
US-10-040-862-8289
; Sequence 8289, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:

```

APPLICANT: Galger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
APPLICANT: Reiter, Marc  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
FILE REFERENCE: 014058-01352005  
CURRENT FILING DATE: 2001-11-06  
PRIOR APPLICATION NUMBER: US/10/040,862  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: US 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: US 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: US 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: US 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: US 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: US 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: US 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: US 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: US 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: US 60/223,378  
PRIOR FILING DATE: 2000-08-07  
PRIOR APPLICATION NUMBER: US 09/796,692  
PRIOR FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 10467  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 8289  
LENGTH: 464  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-040-862-8289

Query Match: 36.2%; Score 414.8; DB 9; Length 464;  
Best Local Similarity 99.5%; Pred. No. 2.1e-96;  
Matches 416; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

729 CTGCCAGTTGGAGCATTGCCCTCTAAGAGCTTTAAACATTTTTCATGCTTATA 788  
DB 1 CTGCCAGTTGGAGCATTGCCCTCTAAGAGCTTTAAACATTTTTCATGCTTATA 60  
789 TACATTCACATTAATGACATTTCTTAATTAATTAACATGATCTGTACTAACA 848  
DB 61 TACATTCACATTAATGACATTTCTTAATTAATTAACATGATCTGTACTAACA 120  
849 CTCACGTGAACCCAGCCTATTGCAAAATTAATCTTTTATATATATCTATGGAT 908  
DB 121 CTCACGTGAACCCAGCCTATTGCAAAATTAATCTTTTATATATATCTATGGAT 180  
909 GTCACACATTAATTAACCTCTGTGGAGAAGTGGAGTTTGGTATTAGTTAATTTTC 968  
DB 181 GTCACACATTAATTAACCTCTGTGGAGAAGTGGAGTTTGGTATTAGTTAATTTTC 240  
969 TAGTAACACAGTTCCTGTTTCTGTTAACAACCTGTATATGCAATTAATATAGGCT 1028  
DB 241 TAGTAACACAGTTCCTGTTTCTGTTAACAACCTGTATATGCAATTAATATAGGCT 300  
1029 TTTCAAATCAGTTAGTAAATAGTACAGATTTAGTTTACATTAAGTCTGACATAC 1088  
DB 301 TTTCAAATCAGTTAGTAAATAGTACAGATTTAGTTTACATTAAGTCTGACATAC 360  
1089 TGGAAATGCATTAAGAGATGTTCAAGTGTCTTTTCAATTTTAAGTAATTTTGT 1146

DB 361 TGGAAATGCATTAAGAGATGTTCAAGTGTCTTTTCAATTTTAAGTAATTTTGT 418

RESULT 6  
US-09-764-864-394  
Sequence 394, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT23  
CURRENT APPLICATION NUMBER: US/09/764,864  
PRIOR APPLICATION data removed - consult PAM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 394  
LENGTH: 864  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (7)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (15)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (9)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (31)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (837)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (853)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (863)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-864-394

Query Match: 22.0%; Score 251.6; DB 10; Length 864;  
Best Local Similarity 72.4%; Pred. No. 2.1e-54;  
Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

152 GGTCTCTGATTTTAACTTATGATATGATGATGATGATGATGATGATGATGATG 221  
DB 79 CATCTGTGATTTTCAAGCTATGATGATGATGATGATGATGATGATGATGATG 138  
222 ATATTACTAGACCAATGATTAAGAGAGAGGAGGACCTAACAAGGCTGATCTG 281  
DB 139 ATGTCTACTACACCAACATGATTAAGGAGGAGGAGGAGGAGGAGGAGGAGG 198  
282 AGCCACCTTAATATGATTAAGTAAATGCTGTTTCAAGGCTGATGATGATGAT 341  
DB 199 CAGCCACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 258  
342 TTGACACATGCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 401  
DB 259 TTGAAATGCTCTGATGATGATGATGATGATGATGATGATGATGATGATG 318  
402 AGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 461  
DB 319 ATGAAGCCACCAACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 378  
462 ATGAACCAACCAACATTTTCTTCAACCAACCAACTCTCTGATGATGATGAT 521

Db 379 AGGGAAACAACAGGAGCTCAACAGAACTCATCTGACCGCCAGCCCTCACCA 438  
Qy 522 ACAATCTGTGTGGAATATGCAAGTATGCTCCGTTTCAAGATGCTGTGAT 581  
Db 439 GCAACACAGTGTGGAATATGCAAGTATGCTCCGTTTCAAGACTGGCGTAGT 498  
Qy 582 AAAGGGCAAAAGTCATCTTCATTTGTC 611  
Db 499 GGGGCTGGCAGAGCTCTTCTTCTTCATTC 528

## RESULT 7

US-09-815-343-1201  
Sequence 1201, Application US/09815343  
Patent No. US20010055596A1  
GENERAL INFORMATION:  
APPLICANT: Meagher, Madeleine  
APPLICANT: Xu, Jiangchun  
APPLICANT: King, Gordon E.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
FILE REFERENCE: 210121.504  
CURRENT APPLICATION NUMBER: US/09/815, 343  
CURRENT FILING DATE: 2001-03-22  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1201  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(318)  
OTHER INFORMATION: n = A,T,C or G  
US-09-815-343-1201

## Query Match

16.1% Score 184.8; DB 10; Length 318;  
Best Local Similarity 73.8%; Pred. No. 1.6e-37;

Matches 234; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 195 CATGTAGAGCTGCTGAGAGTTTGTCAATATTACTATGACACATGATATAAGAAC 254  
Db 1 CCTGACAGAGCTGCTGAGAGTTTGTCAATATTACTATGACACATGATATAAGAAC 60  
Qy 255 GGGCAGCTACAGAGCTGATCTGACACAGCCACCTTAATATGAAATGAAATGCTTT 314  
Db 61 GTTGTCTGCTCCGCTGTGACATGGGACAGCCACCTGTGTGAAATGCAATGCTTT 120  
Qy 315 CAGGCTGAGATGCCCTTAATATTATTTTGAACATTCCTTGAATGAGTCCAGTCA 374  
Db 121 CAGGACAGAAATCCTTGAATGAGTTCCTTGAATGAGTTCCTTGAATGAGTTC 180  
Qy 375 ATATGTAGATTGCTCAACAGTTCATGAGACAGCACTACCTCCAACTACAGTCTTG 434  
Db 181 GCGTGTGAGTACCTGACAGCTTTCATGATGAAGCACCACAGCCAGACAGGCTCTG 240  
Qy 435 TTGTGACCAAGTGAAGTGTGAAGTTTGAAGAAACAAACATTTCTTCAACAGAACT 494  
Db 241 TTGTGATCTGTGATCAGTGAAGTTTGAAGGGAACAAACAGGAGCTTCAACAGAACT 300  
Qy 495 TCCTGCTGACTGCTCAG 511  
Db 301 TCATCTGACCGCCAG 317

## RESULT 8

US-09-796-692-4160  
Sequence 4160, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Galger, Alexander  
APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
FILE REFERENCE: 2077 001200  
CURRENT APPLICATION NUMBER: US/09/796, 692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4160  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-4160

## Query Match

16.1% Score 184.2; DB 9; Length 318;  
Best Local Similarity 73.8%; Pred. No. 2.2e-37;

Matches 234; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 195 CATGTAGAGCTGCTGAGAGTTTGTCAATATTACTATGACACATGATATAAGAAC 254  
Db 1 CCTGACAGAGCTGCTGAGAGTTTGTCAATATTACTATGACACATGATATAAGAAC 60  
Qy 255 GGGCAGCTACAGAGCTGATCTGACACAGCCACCTTAATATGAAATGAAATGCTTT 314  
Db 61 GTTGTCTGCTCCGCTGTGACATGGGACAGCCACCTGTGTGAAATGCAATGCTTT 120  
Qy 315 CAGGCTGAGATGCCCTTAATATTATTTTGAACATTCCTTGAATGAGTCCAGTCA 374  
Db 121 CAGGACAGAAATCCTTGAATGAGTTCCTTGAATGAGTTCCTTGAATGAGTTC 180  
Qy 375 ATATGTAGATTGCTCAACAGTTCATGAGACAGCACTACCTCCAACTACAGTCTTG 434  
Db 181 GCGTGTGAGTACCTGACAGCTTTCATGATGAAGCACCACAGCCAGACAGGCTCTG 240  
Qy 435 TTGTGACCAAGTGAAGTGTGAAGTTTGAAGAAACAAACATTTCTTCAACAGAACT 494  
Db 241 TTGTGATCTGTGATCAGTGAAGTTTGAAGGGAACAAACAGGAGCTTCAACAGAACT 300  
Qy 495 TCCTGCTGACTGCTCAG 511  
Db 301 TCATCTGACCGCCAG 317

## RESULT 9

US-10-040-862-4160  
Sequence 4160, Application US/10040862  
Publication No. US20030078396A1  
GENERAL INFORMATION:  
APPLICANT: Galger, Alexander





RESULT 12  
US-10-239-676-96  
Sequence 96, Application US/10239676  
Publication No. US20030082609A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIOR FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 96  
LENGTH: 15732  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-96

Query Match 3.8%; Score 43.8; DB 9; Length 15732;  
Best Local Similarity 44.5%; Pred. No. 2;  
Matches 174; Conservative 0; Mismatches 217; Indels 0; Gaps 0;  
DB 754 AAGAGCTTTAAACATTTTTCATGCTTATATACATTCACATATGACATTCCTAT 813  
4426 AAGTATATTAATTTAT 111  
DB 814 AATAATATTAACACATGATCTGTGATACATCACTCACTGATGAAAGCCATATGCA 873  
4486 GATATATGTGTATATATAGAGGTTTATTTATAGATGATAGATGAAGATGAAGAG 4545  
DB 874 AATAATATTAACATGATCTGTGATACATCACTCACTGATGAAAGCCATATGCA 873  
4546 GATATATGTGTATATATAGAGGTTTATTTGAAATATTAATATGAAGATGAGTTT 111  
DB 934 AAGAGTGAAGTTTGTGTTATAGTTTATTTTCTAGTAAACACAGTTGCCGTTTCA 993  
4606 TTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4665  
DB 994 GTTAACATGTAATGCAATTTTAAATATATATATATATATATATATATATATATAT 1053  
4666 GTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4725  
DB 1054 TACAGTTTATGTTTAT 1113  
4726 TAGGCTTTTAAATAG 4785  
DB 1114 TGGCTTTTAT 1144  
DB 4786 TATATGTAAGATTTTAAATGTTTATTTT 4816

RESULT 13  
US-10-239-676-90  
Sequence 90, Application US/10239676  
Publication No. US20030082609A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian

APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIOR FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 90  
LENGTH: 5689  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
NAME/KEY: unsure  
LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)  
US-10-239-676-90

Query Match 3.8%; Score 43; DB 9; Length 5689;  
Best Local Similarity 44.5%; Pred. No. 1.8; 243; Indels 6; Gaps 1;  
Matches 200; Conservative 0; Mismatches 243; Indels 6; Gaps 1;  
DB 524 AATACGTGTGAAGATGCAATGCAATGCTTCCGTTTCAATGCTTACTAGTTAA 583  
1610 ATTATTTTGAATGAT 1669  
DB 584 AGGGGCAAAAGTCCATTCATTTGCTTCATTTAGTTCCAGCAATGAATTTATGTAAT 643  
1670 ATGCGATATTTTATTTGATATATATATATATATATATATATATATATATATATAT 1729  
DB 644 TATTTGATTTGAAGACAT 1789  
1730 TTTTAT 1789  
DB 704 TTCCGTGACACCTTTTTCAGACGCGCAGTTTGAGCATGCGCTTAAGAGCTTTA 763  
1790 TTATTTATTTTATTTAT 1843  
DB 764 AATCTATTTTTCATGCTTATATATATATATATATATATATATATATATATATATAT 823  
1844 AATTAATTTAT 1903  
DB 824 AACCATGATCTGTGATACATCACTCACTGAAAGCCATATGCAAAATTAAT 883  
1904 TTAT 1963  
DB 884 CTTTAT 943  
1964 ATTTTAT 2023  
DB 944 TTTTGGTTATATAGTTAATTTTCTAGT 972  
DB 2024 TTTTATTTTATATGTTTATTTTATATAT 2052

RESULT 14  
US-10-239-676-209  
Sequence 209, Application US/10239676  
Publication No. US20030082609A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIOR FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 209  
LENGTH: 11812  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-209

Query Match 3.7%; Score 42.8; DB 9; Length 11812;  
Best Local Similarity 45.6%; Pred. No. 3.1;  
Matches 187; Conservative 0; Mismatches 222; Indels 1; Gaps 1;

QY 633 TTATGCGAATTTTGTGATGAGAGACATATATGCTGGAACATAATTCCTT 692  
DB 4307 TTTATGCTTATGCTTTTGTGACGTTTAAATATATGTTTATTTGTTTGT 4366  
QY 693 AATATTTTCATCCCTGTCGACCTTTCTAGACGTCGACGTTGACATTCCTC 752  
DB 4367 TTAATTTATTTTATTTTATTTATAGTTTATTTATTTATTTATTTATTT 4426  
QY 753 TAAGAGCTTAAACATTTTATTTTACATGCTTATATACATTCCTT 811  
DB 4427 TTTATTTTATTTTATTTTATTTTACGTTTATTTTATTTTATTTATTTAT 4486  
QY 812 ATATATATTTAAACACATGATCTGCTACTACATCATCTGTCGACCGCTAT 871  
DB 4487 TTAATTTTATTTTATTTATTTATGTTTATTTATTTATTTATTTATTTAT 4546  
QY 872 CAAAAATTAATCTTTTATTAATTTATCTATGAGATCGACGACATATACACT 931  
DB 4547 TGTTTAAATTTATTTTATTTTATTAATTCGTTTAAATTTAGCGTATTAAT 4606  
QY 932 GAAGAGCTGAGTTTGTGTTATTTAGTTTATTTCTAGTAAACAGTTGCTGTTT 991  
DB 4607 GAATTAAGTTAGATTTTGTGTTATTTTGTATGTTTATTTGGAAGATGAGG 4666  
QY 992 CAGTTAAGCTGTTATGCTTTTATTAATTTATGCTTTTCAATTCAGTT 1041  
DB 4667 TTGTTTGTGCTATTTTATTAATTTATTTGTTTATTTATTTATTTATTTAT 4716

RESULT 15  
US-10-172-086-18/c  
Sequence 18, Application US/10172086  
Publication No. US20030113750A1  
GENERAL INFORMATION:  
APPLICANT: Epigenomics AG  
TITLE OF INVENTION: Method and nucleic acids for the differentiation  
FILE REFERENCE:  
CURRENT APPLICATION NUMBER: US/10/172,086  
CURRENT FILING DATE: 2002-06-13  
NUMBER OF SEQ ID NOS: 116  
SEQ ID NO 18  
LENGTH: 18997  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

US-10-172-086-18

Query Match 3.7%; Score 42.8; DB 9; Length 18997;  
Best Local Similarity 47.7%; Pred. No. 4;  
Matches 125; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 641 AATTTATTTGATTTGTAAGCACTATATATGTCGTAACATAATTTCTTATATTT 700  
DB 17502 AATTTATTTGATTTGTAAGCACTATATATGTCGTAACATAATTTCTTATATTT 17443  
QY 701 CTATTCCTGTCACACCTTTTCTAGACGTCGACGTTTGAGCATTCGCTGAGAGCT 760  
DB 17442 TACTACTTTTCTAATAAATTAATCAAAATTAATTAATTAATTAATTAATTA 17383  
QY 761 TTAACATATTTTATACATGCTTATATATATTCCTATGATTTCTATATATA 820  
DB 17382 TTACAACTTATTTCTCTTAAATTTTATTTTCAAAATTAATTAATTAATTTCTATATTC 17323  
QY 821 TTAACGATGATCTTGTGCTACATACATCTGATGACCGACCTATGCAAAATAA 880  
DB 17322 TATTTAAACCAAAAAAATTTTATATCACACTTAAACCAAAATTTTAAAAATATA 17263  
QY 881 AATCTTTTATTAATTTATCTA 902  
DB 17262 ATATATATATCAATTAACATTA 17241

Search completed: June 28, 2003, 03:41:59  
Job time: 191.807 secs

GENOME  
RESEARCH



Herculase™

HOME HELP FEEDBACK SUBSCRIPTIONS ARCHIVE SEARCH TABLE OF CONTENTS

Institution: US PATENT & TRADEMARK OFFICE Sign In as Individual

Genome Research, Vol 6, 791-806, Copyright © 1996 by Cold Spring Harbor Laboratory Press

## ARTICLES

# Normalization and subtraction: two approaches to facilitate gene discovery

MF Bonaldo, G Lennon and MB Soares

Department of Psychiatry, College of Physicians and Surgeons of Columbia University, New York, New York, USA.

- ▶ [Email this article to a friend](#)
- ▶ Similar articles found in:  
     [Genome Online](#)  
     [PubMed](#)
- ▶ [PubMed Citation](#)
- ▶ This Article has been cited by:  
     [other online articles](#)
- ▶ Search PubMed for articles by:  
     [Bonaldo, M. F.](#) || [Soares, M. B.](#)
- ▶ Alert me when:  
     [new articles cite this article](#)
- ▶ [Download to Citation Manager](#)

Large-scale sequencing of cDNAs randomly picked from libraries has proven to be a very powerful approach to discover (putatively) expressed sequences that, in turn, once mapped, may greatly expedite the process involved in the identification and cloning of human disease genes. However, the integrity of the data and the pace at which novel sequences can be identified depends to a great extent on the cDNA libraries that are used. Because altogether, in a typical cell, the mRNAs of the prevalent and intermediate frequency classes comprise as much as 50-65% of the total mRNA mass, but represent no more than 1000- 2000 different mRNAs, redundant identification of mRNAs of these two frequency classes is destined to become overwhelming relatively early in any such random gene discovery programs, thus seriously compromising their cost-effectiveness. With the goal of facilitating such efforts, previously we developed a method to construct directionally cloned normalized cDNA libraries and applied it to generate infant brain (INIB) and fetal liver/spleen (INFLS) libraries, from which a total of 45,192 and 86,088 expressed sequence tags, respectively, have been derived. While improving the representation of the longest cDNAs in our libraries, we developed three additional methods to normalize cDNA libraries and generated over 35 libraries, most of which have been contributed to our integrated Molecular Analysis of Genomes and Their Expression (IMAGE) Consortium and thus distributed widely and used for sequencing and mapping. In an attempt to facilitate the process of gene discovery further, we have also developed a subtractive hybridization approach designed specifically to eliminate (or reduce significantly the representation of) large pools of arrayed and (mostly) sequenced clones from normalized libraries yet to be (or just partly) surveyed. Here we present a detailed description and a comparative analysis of four methods that we developed and used to generate normalized cDNA libraries from human (15), mouse (3), rat (2), as well as the parasite *Schistosoma mansoni* (1). In addition, we describe the construction and preliminary characterization of a subtracted liver/spleen library (INFLS-SI) that resulted from the elimination (or reduction of representation) of ~5000 INFLS-IMAGE clones from the INFLS library.

NCBI Nucleotide

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Bio

Search  for

☐ 1: AW274482.xv30c09.x1 Soares...[gi:6661512]

[Links](#)

#### IDENTIFIERS

dbEST Id: 3657583  
EST name: xv30c09.x1  
GenBank Acc: AW274482  
GenBank gi: 6661512

#### CLONE INFO

Clone Id: IMAGE:2814640 (3')  
Source: NCI  
DNA type: cDNA

#### PRIMERS

Sequencing: -40UP from Gibco  
PolyA Tail: Unknown

#### SEQUENCE

```
AGAAAAATAGAAAAGTTTACTACTTTGAAAAGGAACTATGACAAACAAGTATATATTCA
GGAAAGGGACTCCTAGAACTTGAGCAACAAAAAGAGTTCAACCTCGGCACAGTGGAGTC
TGCAGTGTCTCCGCAACAACAGAAATGAAGTTGTGTTCTCACATCCTTGAGAGTCGAGG
TTTGCATTTCTCTACAGAGCTAGGGCTGAATGAAGCAAAGAGACCTCTGCCACCCCACT
AGCTGGCCCAGTCCTGGAAGCGGAAGCAGTCACTTGGCATCTTCCACACTGTGTTGCTGG
GTGAGGCCTGGGCGGTGAGGATGAAGTTCTGGTTGAAGTCCCGTTGTTTGGTCCCCTCAA
ACTTCACTGATCCACAGATGACAACAAGGACCGTGGTCTGGCTTGGTGTGGCTTCATCAT
GAACAGGCTGGCAGTCTACCACGCTGATTTGGAAGTCTGGTGAAGGCAACATTTCAAAA
ACTCACTCAAGGATTCTTGTCTGAAACAGCATTGCCATTCCAGACCAGGTGGCTGTGC
CCATGTACAGGCGGGACAGCAAACGCCGCGCTTATCCATGGTGGTGTAGTAGACATTGA
CAGACTCCTCAGCAGCTCTGCAGGCCTGATC
```

Quality: High quality sequence stops at base: 466

Entry Created: Jan 3 2000  
Last Updated: Jan 3 2000

#### COMMENTS

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

PUTATIVE ID Assigned by submitter  
WP:Y71F9A\_279.B CE22843 ;

#### LIBRARY

Lib Name: Soares\_NFL\_T\_GBC\_S1  
Organism: Homo sapiens  
Organ: pooled  
Lab host: DH10B  
Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker

h      cb      hg   e   e   e   fcg   c      e   e   e   b   c   e   e

R. Site 1: Not I  
R. Site 2: Eco RI  
Description: Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI\_CGAP\_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

**SUBMITTER**

Name: Robert Strausberg, Ph.D.  
E-mail: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

**CITATIONS**

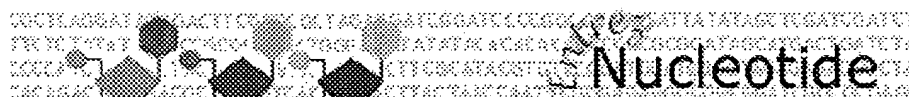
Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Authors: NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>  
Year: 1997  
Status: Unpublished

**MAP DATA**

---

[Disclaimer](#) | [Write to the Help Desk](#)  
[NCBI](#) | [NLM](#) | [NIH](#)

10/30/2003 12:44:50



PubMed

Nucleotide

Protein

Genome

Structure

PMC

Taxonomy

OMIM

Bio

Search  for   

Limits

Preview/Index

History

Clipboard

Details

Display

Show:

Send to

☐ 1: BF195575. 7n85b11.x1 NCI\_CG...[gi:11082605]

Links

## IDENTIFIERS

dbEST Id: 6660525  
EST name: 7n85b11.x1  
GenBank Acc: BF195575  
GenBank gi: 11082605

## CLONE INFO

Clone Id: IMAGE:3571173 (3')  
Source: NCI  
DNA type: cDNA

## PRIMERS

PolyA Tail: Unknown

## SEQUENCE

AAGAAGCAACATGTGTATTATTAAATTATTTGTCAGAATTTCCAGAATCAGAGTCTCTAC  
TGGGCAAGTAGAAAAATAGAAAAGTTTACTACTTTGAAAAGGAACTATGACAAACAAGT  
ATATATTTCAGGAAAGGGACTCCTAGAACTTGAGCAACAAAAAGAGTTCAACCTCGGCAC  
AGTGGAGTCTGCAGTGTCTCCGCAACAACAGAAATGAAGTGTGTTCCCTCACATCCTTGA  
GAGTCGAGGTTTGCATTTCTCTACAGAGCTAGGGCTGAATGAAGCAAAGAGACCTCTGCC  
ACCCCCACTAGCTGGCCAGTCCTGGAAGCGGAAGCAGTCACTTGCATCTTCCACACTG  
TGTTGCTGGGTGAGGCCTGNGCGGTGAGGATGAAGTCTGGTTGAAGTCCCCTTGTGTG  
TCCCCTCAAACCTTCACTGATCCACAGATGACAACAAGGACCGTGGTCTGGCTTGGTGTGG  
CTTCATCATGAACAGGCCTGCAGTCTACCACGCTGATTTGGAAGTCCGCTGGAAGGCAACA  
TTTCAAAAACTCACTCAAGGATTCTTGTGCTGAAACAGCATTGCCATTCAGAACCAAGG  
GGTGCTGTGCCCATGTACAGGCCGGACAGCAAACGCCCCGCGCTTATCCCTGGTGGTGTG  
TGACATTGACAACTTCTCACAGCTCTGCAAGGCTGATCCACATAGTTGTGAAATCCCAA

Quality: High quality sequence stops at base: 466

Entry Created: Nov 3 2000  
Last Updated: Nov 3 2000

## COMMENTS

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D.,  
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library  
Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by:  
Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by:  
Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information  
can be found through the I.M.A.G.E. Consortium/LLNL, send  
email to: [info@image.llnl.gov](mailto:info@image.llnl.gov)

## PUTATIVE ID

Assigned by submitter  
TR:Q9UKK6 Q9UKK6 NTF2-RELATED EXPORT PROTEIN NXT1..;

## LIBRARY

h cb hg e e e fcg c e e e b c e e

Lib Name: NCI\_CGAP\_Ov18  
Organism: Homo sapiens  
Organ: ovary  
Tissue type: fibrotheoma  
Lab host: DH10B (phage-resistant)  
Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker  
R. Site 1: Not I  
R. Site 2: Eco RI  
Description: 1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGACATTTTTTTTTTTTTTTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library went  
through one round of normalization, and was constructed by  
Bento Soares and M. Fatima Bonaldo.

**SUBMITTER**

Name: Robert Strausberg, Ph.D.  
E-mail: cgapbs-r@mail.nih.gov

**CITATIONS**

Title: National Cancer Institute, Cancer Genome Anatomy Project  
(CGAP), Tumor Gene Index  
Authors: NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>  
Year: 1997  
Status: Unpublished

**MAP DATA**  
=====

[Disclaimer](#) | [Write to the Help Desk](#)  
[NCBI](#) | [NLM](#) | [NIH](#)

Jul 30 2003 12:44:50